情報処理学会論文誌　データベース　Vol. 3　No. 2　61–79 (June 2010)

データ対間のサポート関係分析に基づくWeb情報の信憑性評価

山本祐輔[†1,†2]　田中克己[†1]

本稿では，ウェブ情報の信憑性を分析するための汎用的なモデルを提案する．提案モデルでは，任意の観点からの信憑性分析を行うために分析対象となるウェブ情報とその関連情報を何らかのデータの対で表現する．そのうえで，データ対間の support 関係の強さを分析することで，対象となっているウェブ情報の信憑性を評価する．提案モデルではデータ対間の support 関係の強さを定義するために，2 種類の観点を導入する．提案モデルでは，「supportive なデータ対を多く持つデータ対は信憑性が高い」という仮説のもと，データ対として表現されたウェブ情報の信憑性分析を行う．提案モデルは汎用的であり，信憑性の分析対象となっている情報の種類に応じて support 関係の設定を行うことで，様々なウェブ情報の信憑性を分析することが可能となる．本稿では，例題として Q&A サイト上の回答の信憑性を評価するために質問・回答対間の support 関係を分析するケース，および画像信憑性を評価するために画像・テキスト対間の support 関係を分析するケースを取り上げ，実際に提案モデルによる信憑性分析を行った．さらに，ウェブ情報の真偽判定に対する提案モデルの有効性を評価する実験を行った．評価実験では，分析によって得られた信憑性スコアを用いることで，Q&A サイト上の回答の真偽を 70%以上の精度で，画像の真偽を 65%以上の精度で判定することができた．

Modeling and Measuring Web Information Credibility by Analyzing Support Relation between Data-pairs

Yusuke Yamamoto[†1,†2] and Katsumi Tanaka[†1]

In this paper, we propose a model for analyzing the credibility of Web information. In our model, target Web information and its related information are represented as data-pairs. We analyze the credibility of a target Web information that is represented as a data-pair by checking the supportive relationship between the target data-pair and its related data-pairs. We introduce two types of models for data-pair relationships. Our hypothesis is that the more supportive data-pairs a data-pair has, the more credible it is. Our model is generic, so it can be applied to a variety of types of Web information. In this paper, we selected some data-pairs such as question-answer pairs on Q&A sites to evaluate the credibility of the answers and image-text pairs to evaluate the credibility of images, and analyzed the credibility of them using our proposed model. Moreover, we conducted experimental evaluation to reveal how useful our credibility analysis is for judging the correctness of Web information. The experimental results showed that the system automatically judged the correctness of question-answer pairs with accuracy of over 70% and the correctness of image-text pairs with accuracy of over 65% by using calculated credibility scores.

1. はじめに

ウェブの存在が日常生活と切り離せないほど重要になったことで，ウェブ情報の信憑性を担保することが重要な研究課題として認知されはじめている．ユーザはウェブ上で自由に情報を閲覧，発信することができるが，既存のマスメディアのようにウェブ情報の質が発信の前段階で厳しく評価されることはない．それゆえ，ウェブ情報のすべてが正確であるとは限らない．たとえば，ウェブには 20000 以上の医療情報サイトが存在し，そのようなサイトの半数以上が医療専門家によるチェックを受けていないとの報告もなされている[33]．さらに，昨今最も利用されているウェブコンテンツの 1 つである Wikipedia[⋆1]の信憑性に対しても，Denning らによって疑問が投げかけられている[9]．一方で，多くのユーザが，ウェブに存在する情報はある程度信用できるものとして認知しているとの研究報告もなされている[27),28)]．ユーザがウェブ情報の信憑性を意識せずに情報を閲覧してしまった場合，誤った情報を鵜呑みにしてしまい，場合によっては実害を被る可能性も十分にある．

オンライン情報の信憑性確保に関しては図書館学研究のコミュニティでは以前から重要な研究課題として認知されており[25),27)]，オンライン情報の信憑性を評価に関するガイドラインの策定や情報の信憑性に影響を与える要因の解明に研究の重点が置かれてきた．心理学や社会学の分野では，古くから信憑性が重要なトピックとして研究対象とされてきた[13),14),24),34),37)]．これらの分野では，情報の信憑性は「ユーザによって “認知” される情報の正しさの度合い」

†1 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University
†2 日本学術振興会特別研究員（DC2）
a Research Fellow (DC2) of the Japan Society for the Promotion of Science
⋆1 Wikipedia. http://wikipedia.org/

として定義されている．すなわち，**信憑性とはある視点から推定される情報の正しさ**を意味する．それゆえ，同じ情報でも視点によって信憑性の高低は異なり，また信憑性の高い情報が実際には誤りである可能性があると解釈される．情報の信憑性に影響する要素は多数あげられているが，大きくは trustworthiness と expertise の 2 つの要素に集約される[14),34)]．Expertise は情報発信者が有する情報記述能力に関する要素で，読み手が情報を閲覧した際に認知される発信者の知識量，経験，技量の程度を意味する．Trustworthiness は，読み手によって認知される情報発信者の道徳性や優良性に関する要素で，情報の公正さや偏見の有無，評判の良さなどに関係がある．

このように，図書館学研究者によって信憑性検証のガイドラインの整備が，心理学者や社会学者による情報信憑性の定義，信憑性に影響を与えている要因の考察がなされてきた．しかし，このような努力にもかかわらず，現状では，ユーザにとってオンライン情報の信憑性を適切に評価することは困難である．これは，従来のメディアコンテンツと比べオンライン情報，特にウェブ情報は，発信者の名前，肩書きなど発信者の能力を評価するための情報が乏しい，扱われている情報のソース，証拠情報が特定しにくい，情報の発信日時が不明確である，といったように信憑性を推定する情報が乏しいことが原因としてあげられる．その結果，ユーザの多くは信憑性検証を適切に行うことができない．このような状況を解決するためにも，ウェブ情報の信憑性を評価・判断支援するためのシステムが必要となる．

ウェブ情報の評価方法に関する研究は情報検索の分野でさかんに行われてきた．しかし，そのような研究の多くは，クエリとの適合性評価やリンク解析に基づく人気度評価[7),8),17)]，クリックスルーデータを用いた人気度評価[22)]に基づくデータセットのランキングに焦点が当てられており，信憑性に関する評価手法が提案されることはなかった．近年では，情報の質に関する議論も始まりつつあり，我々の研究グループも含め，情報の質の一側面として信憑性の重要性を主張する研究者も現れている[31)]．しかし，情報の質に関する研究は始まったばかりで，特にウェブ情報の信憑性に関して具体的に評価手法を検討している研究事例は依然少ない．

本稿の目的は，情報検索，心理学，社会学を横断し信憑性評価手法を議論することである．本稿では，ウェブ情報の信憑性を評価するための汎用的なモデルを提案する．さらに，提案する信憑性分析モデルがウェブ情報の実際の真偽を判断するうえでどの程度有効に機能するかについての議論も行う．提案モデルでは，分析対象となるウェブ情報と信憑性の分析観点が与えられたとき，分析対象のウェブ情報とそれに関連するウェブ情報をデータ対で表現し，データ対間の関係性の強さを考慮することで，対象となっているウェブ情報の信憑性を分析する．たとえば，質問応答サイトにおける回答の信憑性を分析するケースでは質問・回答対，画像の信憑性を分析するケースでは画像・テキスト対などが信憑性分析を行う際に着目するデータ対の例としてあげられる．

対象となっているウェブ情報の信憑性をそれ単体のみから分析することは非常に困難である．Meola[25)]が述べるように，ある情報の信憑性を評価する有力な方法の 1 つは，関連する別の情報との比較を繰り返すことである．対象となっている情報が他の情報ソースでも言及されていることを確認することは，trustworthiness の観点から信憑性を確認するという意味で非常に重要である．実際，ユーザはこの種のアプローチをとることが多い．Metzger や Rieh らの調査によると，ユーザの多くはウェブ情報の信憑性を確認するために，関連するウェブページを検索し，信憑性分析の対象となっているウェブページの内容と比較する，との報告がなされている[26),32)]．本稿で提案するアプローチも同様のアプローチである．提案モデルでは，データ対として表現されたウェブ情報間の support 関係に着目する．提案モデルでは，信憑性分析対象であるデータ対に対して “supportive” なデータ対が多く存在するほど，対象データ対の信憑性が高まる，という仮定を置く．“Supportive” の定義は対象としているウェブ情報の種類とそれが用いられるコンテキストに依存する．提案モデルでは汎用的に構成されており，信憑性分析の対象に応じて support の定義を行うことで様々なウェブ情報の信憑性を分析することが可能となる．以下に本稿の新規性，進歩性を記す：

- 信憑性分析観点に基づき対象ウェブ情報と関連情報をデータ対で表現し，データ対間の “support” 関係を分析することでウェブ情報の信憑性を評価する汎用的なモデルを提案した点
- 提案モデルを適用例として質問応答サイトに投稿された回答の信憑性およびウェブページに掲載された画像の信憑性を例題として取り上げ，信憑性分析のための具体的な実装方法を示した点
- 提案モデルによる信憑性分析結果とウェブ情報の実際の真偽との関係を分析した点

2. データ間の関係に基づく信憑性分析

本稿では，特定の信憑性観点からウェブ情報の信憑性を分析するための汎用的なモデルについて述べる．ウェブ情報の信憑性を特定の観点から信憑性を分析することは，分析対象であるウェブ情報のデータと別種のデータの関係性を分析することに相当すると考えることができる．たとえば，質問応答サイト（以下 QA サイト）に投稿されている回答の信憑性を分析する場合，信憑性の観点の 1 つとして質問に対する回答の典型性などが考えられるが，

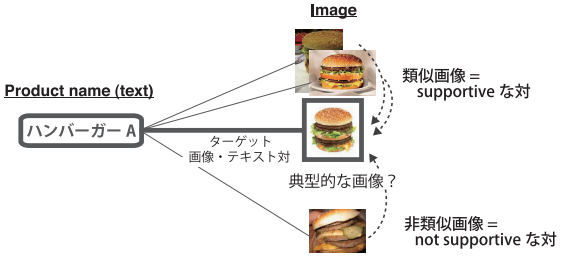


図 1　分析対象の画像・テキスト対とそれと関連する画像・テキスト対の関係
Fig. 1　Relation between target image-text pair and its related pairs.

これは質問と回答のデータ間の関係性を分析することに相当する．そこで，本章で提案する信憑性分析では，**特定の信憑性観点からウェブ情報の信憑性分析を行うために，対象となっているウェブ情報を何らかのデータの対で表現し他の関連データ対との関係を分析する．すなわち，特定観点からのウェブ情報の信憑性分析をデータ対の信憑性分析と見なし分析を行う**．

あるデータ対で表現されるウェブ情報の信憑性を評価するために，提案モデルでは，まず分析対象のデータ対と同じデータ対で表現され，かつ対象となっているウェブ情報と関連するデータ対を収集する．その後，収集された各データ対が分析対象のデータ対に対してどの程度 “supportive” かを分析する．最終的に分析対象のデータ対がどの程度 supportive なデータ対を持つかを分析することでデータ対の信憑性を評価する．

“Supportive” の定義は対象としているウェブ情報の種類とウェブ情報が使用されているコンテキストに依存する．一例として**図 1** に画像信憑性を分析するために画像・テキスト対に注目し，分析対象となっている画像・テキスト対および関連する画像・テキスト対の関係を検証する例を記す．今，太線で囲まれた画像がある商品の様子を示す典型的な画像であるかを疑っている状況を想定する．すなわち，この例ではある商品 X に対する画像典型性を画像信憑性の基準として着目している．このケースにおいて，信憑性分析対象（ターゲット）のデータ対に対して “supportive” なデータ対とは，同じ商品に関する画像のうちターゲットとなっている画像と類似するものを指す．提案モデルを用いてターゲット画像の信憑性を評価する場合，ターゲット画像と類似するハンバーガー A に関する画像がウェブ上にどの程度存在するかを分析する．図 1 では，ターゲットである画像・テキスト対はハンバーガー A に関して類似する画像を多く持つことから，関連データ対から多くの support を受けていると考え，結果的にターゲット画像・テキスト対は信憑性が高い，すなわちテキストに対する画像の典型性という観点から画像の信憑性は高いと評価される．

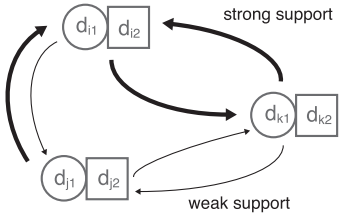


図 2　データ対間の関係を表すグラフ．線の太さはデータ対ノード間の support の強さを示している
Fig. 2　Graph representation of relationship between data-pairs. Line weight represents the strength of support between data-pair nodes.

このように，本稿で提案する信憑性分析モデルでは，“supportive” の概念を導入したうえで，**多くの信憑性の高いデータ対がターゲットのデータ対に対して supportive であるほど，ターゲットデータ対の信憑性は高い**，という基本的な仮説のもとで信憑性分析を行う．$sup(p_i, p_j)$ をデータ対 p_j のデータ対 p_i に対する support の大きさとしたとき，先の仮説に基づき信憑性スコア $cred(p_i)$ を以下のような式で再帰的に定義することができる：

$$cred(p_i) = \sum_{p_j \in RelatedPair(p_i)} sup(p_i, p_j) \cdot cred(p_j) \quad (1)$$

ここで，$RelatedPair(p_i)$ はデータ対 p_i に関連するデータ対の集合を意味する．$sup(p_i, p_j)$ と $cred(p_j)$ の積は，データ対 p_i の信憑性はデータ対 p_j 自身の信憑性とデータ対 p_i への support の大きさの両方の要素によって決まることを表している．

データ対間の support 関係を定義すると，各データ対間の関係は**図 2** のように各データ対をノード，データ対間の support 関係を重み付きのエッジと見なしたグラフによって表現することができる．これにより，式 (1) による信憑性スコアの計算は，LexRank [11] や VisualRank [18] のようなグラフの中心性の評価する問題に帰着することができる．

実際に信憑性分析のターゲットとなっているデータ対およびその関連データ対の信憑性スコアを計算するために，全データ対間の support の大きさと各データ対の信憑性スコアを行列で表現する．このとき，式 (1) は行列表現を用いて以下のように再定義される：

$$\boldsymbol{CS} = \boldsymbol{S}^* \times \boldsymbol{CS} \quad (2)$$

ここで，$\boldsymbol{S}^*$ は i 行 j 列目の要素 $\boldsymbol{S_{i,j}}$ がデータ対 p_j のデータ対 p_i に対する support の大きさを表している行列 $\boldsymbol{S}$ を各列ごとに正規化した行列，$\boldsymbol{CS}$ は第 i 番目の要素 $\boldsymbol{CS_i}$ がデー

タ対 p_i の信憑性スコア $cred(p_i)$ を表すベクトルを意味する．最終的に，提案モデルでは行列 $\boldsymbol{S}^*$ の固有値を求めることで各データ対の信憑性スコアを計算することができる．

提案モデルでは，データ対の種類およびそれが現れるコンテキストに応じてデータ対間の support の大きさ sup を定義することで，様々な信憑性を柔軟に評価することができる．データ対の信憑性評価に複数の信憑性分析の観点を導入したい場合は，support の大きさ sup をうまく定義することで対応することも可能である．

2.1　信憑性分析の観点に応じた support の定義

Support の定義に関する議論を行うために，2 つのデータ対 $p_i = (d_{i1}, d_{i2})$，$p_j = (d_{j1}, d_{j2})$ を取り上げる．ここで，d_{i1}，d_{j1}，d_{i2}，d_{j2} はデータ対 p_i，p_j の構成要素であり，データ d_{i1} と d_{j1}，d_{i2} と d_{j2} はそれぞれ同じデータ型である．データ対 p_j の p_i に対する support 値 $sup(p_i, p_j)$ はデータ d_{i1} と d_{j1} 間の距離 $dist_1(d_{i1}, d_{j1})$ およびデータ d_{i2} と d_{j2} 間の距離 $dist_2(d_{i2}, d_{j2})$ を用いて表現する．Support の定義はすでに述べたようにウェブ情報の信憑性の解釈に応じて異なる．そこで以下では，正規化したデータ間の距離 $dist_1$ および $dist_2$ の組合せを考慮することで，データ対間の関係分析によって分析可能な 2 種類の基本的な信憑性について論じる．

Dominance

基本となる信憑性観点の 1 つ目は関連データ対集合内における対象データ対の dominance である．Dominance はターゲットとなっているデータ対が関連データ対集合中で類似するデータ対をどの程度持つかを意味する．たとえば，図 1 では画像の信憑性の 1 つの観点として「ある商品に関してより多くの画像がターゲット画像と類似しているほどターゲット画像は信憑性は高い」と考えられる．これは画像の信憑性を分析するために画像・商品名のデータ対に注目し，dominance に着目したデータ対の信憑性分析と考えることができる．Dominance を信憑性観点として用いる場合，あるデータ対がターゲットデータ対と類似しているほどターゲットデータ対は大きな support を得ることになる．このケースでは，**図 3** の (a) のように，データ対間の類似度はデータ対を構成しているデータノード間の距離 $dist_1$ および $dist_2$ が小さいほど大きくなる．図 1 の例では，$dist_1$ はテキスト（商品名）間の非類似度，$dist_2$ は画像間の非類似度となる．データ対の dominance を評価する場合の support 値は，たとえば直感的には以下のように定義することができる．

$$sup_{dom}(p_i, p_j) = (1 - dist_1(d_{i1}, d_{j1})) \cdot (1 - dist_2(d_{i2}, d_{j2})) \quad (3)$$

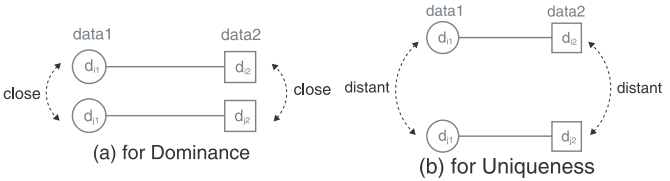


図 3　信憑性分析の観点に応じた support 値の考え方
Fig. 3　Support degree according to credibility aspect.

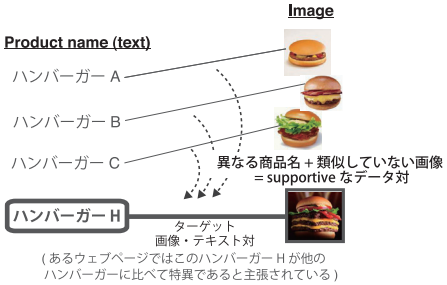


図 4　Uniqueness が信憑性分析の観点として重要になるケース
Fig. 4　Case where uniqueness is important as credibility aspect.

Uniqueness

基本となる信憑性観点の 2 つ目は関連データ対集合内における対象データ対の uniqueness である．Uniqueness はターゲットとなっているデータ対が関連データ対集合内で独立している程度を意味する．この指標は**図 4** のようなケースの信憑性を扱う場合に有効である．今，図 4 では，ハンバーガー H が他のハンバーガーと比べて特異であることがテキスト中で主張されており，それに対する証拠として画像が用いられていると想定する．このケースは特異性を訴えかける主張に対するハンバーガー H の画像の信憑性の問題としてとらえることができる．この場合，ターゲット画像が他のハンバーガー商品の画像と比較して類似性が低いことが，ハンバーガー H が特異であると主張するうえで 1 つの重要な観点となる．これは画像の信憑性を分析するために画像・商品名のデータ対に注目し，uniqueness に着目したデータ対の信憑性分析としてとらえることができる．すなわち，Uniqueness が高いほど先の条件が満たされると考えることができる．図 3 の (b) のように，データ対間の独立度はデータ対を構成しているデー

タノード間の距離 $dist_1$ および $dist_2$ が大きいほど大きくなる．信憑性の観点として uniqueness を考慮する場合，あるデータ対が背反するデータ対を多く持つとき，そのデータ対は support を多く集めており，関連データ対集合中で独立していると考えることができる．データ対の uniqueness を評価する場合の support 値は，たとえば直感的には以下のように定義することができる．

$$sup_{uni}(p_i, p_j) = dist_1(d_{i1}, d_{j1}) \cdot dist_2(d_{i2}, d_{j2}) \quad (4)$$

上で述べた 2 つの信憑性観点はデータ対間の関係性分析に基づくデータ対の信憑性分析を行ううえで基本的な観点となる．データ対の種類や信憑性分析のコンテキストによっては，複数の観点を同時に考慮する必要がある．そのようなケースでは，support 値 $sup(p_i, p_j)$ を以下のように構成することで対応することができる：

$$sup(p_i, p_j) = \alpha \cdot sup_{dom}(p_i, p_j) + \beta \cdot sup_{uni}(p_i, p_j) \quad (5)$$

ここで α，β はそれぞれ重みパラメータであり，$\alpha + \beta = 1$ を満たす．これらのパラメータを調節することで提案モデルを崩すことなく複数の信憑性観点を同時に考慮することができる．たとえば，dominance と uniqueness を同時に同程度考慮して support 値を決定したい場合は，α および β をそれぞれ 0.5 に設定すればよい．

2.2　分布を考慮した信憑性スコアの修正

2.1 節で提案したモデルを使用した場合，ターゲットのデータ対および関連するデータ対の信憑性スコアのリストが得られる．データ対を信憑性スコア順にランキングすることが目的の場合，得られたスコアを直接使用しても問題は生じない．しかし，本研究の目標はデータ対集合のランキングではなく，指定されたデータ対の信憑性を分析することである．関連するデータ対集合中の信憑性スコアの分布を考慮しなければ，ターゲットのデータ対の信憑性を適切に判断することはできない．

たとえば，$\boldsymbol{CS_{raw1}} = (\underline{0.1}, 0.001, 0.001, \ldots, 0.001)^T$ と $\boldsymbol{CS_{raw2}} = (\underline{0.1}, 0.1, \ldots, 0.1)^T$ の 2 種類の信憑性スコアのリストを考える．各リスト中の下線が引かれた 0.1 に着目したとき，$\boldsymbol{CS_{raw1}}$ 中の 0.1 はリスト中の他のスコアと比べて際立って大きい．一方，$\boldsymbol{CS_{raw2}}$ では下線が引かれたスコア以外も同じ 0.1 というスコアを持っており，下線部のスコアを持つデータ対は他のデータ対と比較して信憑性の優劣がつけ難い．つまり，このデータ対は信憑性が高いとはいえない．この例では，同じ信憑性スコアでも $\boldsymbol{CS_{raw1}}$ における信憑性スコア 0.1 の方が $\boldsymbol{CS_{raw2}}$ の信憑性スコア 0.1 よりも信憑性が高いと見なすべきである．それゆえ，式 (2) で得られた全データ対の信憑性スコアの分布を考慮したうえで信憑性スコアを修正する必要がある．

この問題に対応するために，得られた信憑性スコアの分布はガウス分布に従うとの仮定のもと，信憑性スコアを修正する．式 (2) によって得られた信憑性スコアのリストを $\boldsymbol{CS_{raw}} = (c_1, c_2, \ldots, c_n)^T$，$\boldsymbol{CS_{raw}}$ の平均，分散をそれぞれ μ，σ としたとき，元の信憑性スコア c_k の修正スコア c'_k を以下のように定義する：

$$c'_k = \frac{10(c_k - \mu)}{\sigma} + 50 \quad (6)$$

3.　質問応答サイトにおける回答コンテンツの信憑性

近年，質問応答サイト（以下 Q&A サイト）の利用が急速に広がっている．Q&A サイトでは，ユーザが質問を自然言語文で投稿すると，質問に関して知識を有する別のユーザが回答を投稿するという形式をとっている．最も成功している Q&A サイトの 1 つである Yahoo! Answers[⋆1]では，2 億 3000 万以上の質問がすでに投稿されており[1)]，2009 年現在，1 日あたり 90000 の質問が投稿されている[16)]．ウェブ検索エンジンを用いても求めている回答を発見することが困難な場合，このような Q&A サイトは非常に有効である．

しかし，Q&A サイトで投稿される回答の正しさは保証されていない．Q&A サイトでは，ユーザは自身の知識の範囲内で回答可能と “自己判断” した質問に対して自由に回答を投稿できるが，投稿された回答の正しさを検証する機構は Q&A サイトには存在しない．また多くの Q&A サイトは，ユーザが投稿された回答に評価スコアを与えたり，ベストアンサを選出する機能を提供しているが，そのような評価はユーザの主観に依存しており，良い評価を受けた回答が正しい回答であるとは限らない．

本章では，前章で述べたモデルを Q&A サイトに投稿された回答の信憑性分析に適用する手法について述べる．さらに，提案モデルによって評価された回答の信憑性と実際の真偽との関係を評価実験を通して明らかにする．

3.1　アプローチ

Q&A サイト上のある質問に対して正しい回答というものが存在するならば，**図 5** のように，類似質問が Q&A サイトに投稿された場合もそれに対して同じ（正しい）回答が投稿されることが期待される．このような事実から，本稿では回答の信憑性を分析するために**質問・回答（QA）対の信憑性**に着目して以下のような 2 つの仮説を設ける：

- 信憑性の分析対象である QA 対（ターゲット QA 対）が与えられたとき，QA サイト内

⋆1 Yahoo! Answers: http://answers.yahoo.com/

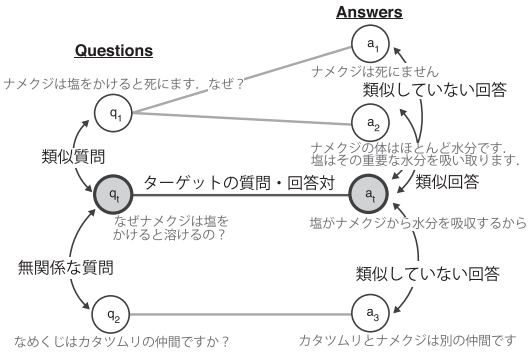


図 5　ターゲットの質問・回答対と他の質問・回答対との関係
Fig. 5　Relation between target question-answer pair and its related pairs.

に存在する質問 X がターゲット対の質問に類似し，かつ質問 X に対して投稿された回答 Y がターゲット対の回答と類似している場合，ターゲット QA 対の信憑性は高まる．

- このような対がターゲット QA 対に対して多数存在すれば，ターゲット QA 対の信憑性は高まる．

2 章で述べたモデルを QA 対の信憑性分析に適用する場合，上記の仮説は **dominance モデル**を用いて定式化することができる．今，2 つの QA 対 $p_1 = (q_1, a_1)$ と $p_2 = (q_2, a_2)$ が与えられたとき，質問 q_1，q_2 間の距離 $dist_1(q_1, q_2)$ および回答 a_1，a_2 間の距離 $dist_2(a_1, a_2)$ を以下のように定義する．

$$dist_1(q_1, q_2) = 1 - sim_t(v_{q_1}, v_{q_2}) \tag{7}$$

$$dist_2(a_1, a_2) = 1 - sim_t(v_{a_1}, v_{a_2}) \tag{8}$$

ここで，v_q と v_a はそれぞれ質問 q と回答 a の特徴ベクトルを表し，$sim_t(v_1, v_2)$ はベクトル v_1 と v_2 のコサイン類似度を表す．このとき，support 値 $sup(p_1, p_2)$ は式 (3) を用いて以下のように定義できる：

$$sup(p_1, p_2) = sup_{dom}(p_1, p_2) \tag{9}$$

$$= sim_t(v_{q_1}, v_{q_2}) \cdot sim_t(v_{a_1}, v_{a_2}) \tag{10}$$

定義された support 値と式 (2)，(6) を用いることで QA 対の信憑性スコアを計算することができる．以下に QA 対の信憑性スコアの計算フローを示す：

(1)　ターゲット QA 対をシステムに入力する．
(2)　ターゲット QA 対の質問から関連する QA 対を収集するためのクエリを抽出する．
(3)　抽出したクエリを用いて QA サイトから関連する QA ペアを検索・収集する．
(4)　ターゲット QA 対および収集された QA 対を構成している質問および回答の特徴ベクトルを生成する．
(5)　ターゲット QA 対の質問と収集された QA 対の質問間の類似度を計算することで，ターゲット QA 対の質問と類似した質問を持つ QA 対を絞り込む．
(6)　ターゲット QA 対と絞り込まれた QA 対の集合に対して式 (2)，(6)，(10) を適用することで，ターゲット QA 対の信憑性スコアを計算する．

3.2　評　　価

提案モデルを用いた QA 対の信憑性分析を評価するために実データを用いて実験を行った．実験の目的は提案手法によって算出された QA 対の信憑性スコアと質問に対する回答の実際の真偽との関係を明らかにすることである．

Q&A サイトの回答に正誤フラグが付与されたテストセットは我々の知る限り存在しない．そこで，テストセットを作成するために，国内で最も利用されている QA サイトである Yahoo! 知恵袋[⋆1]からランダムに質問を抽出し，質問に対する回答の正誤判定が可能なファクト型の質問とその回答をテストセットに加えた．質問に対する回答文の正誤の判定は，各質問に関する Wikipedia 記事，専門家のページなどを複数ページ閲覧し，回答文に関する記述が存在すれば正解，回答文の内容が誤りであるとの記述が存在すれば誤りと判定した．このような作業で判断ができないものに関しては「どちらでもない」と判定した．最終的に計 40 個の質問を選択し，各質問に対して投稿された計 200 個の回答に「正しい（true）」「誤っている（false）」「どちらとも判断できない（neutral）」のいずれかのラベルを付与した．True ラベル，false ラベル，neutral ラベルの個数はそれぞれ 114 個，63 個，23 個であった．**表 1** にテストセットの例を記す．

本実験では，信憑性分析のために関連 QA 対を収集する段階で必要なクエリは著者が手動で指定した．検索時に使用したクエリに関しては表 1 に下線で記した．関連 QA 対の収集には Yahoo! 知恵袋・質問検索 API [⋆2]を使用した．特徴ベクトル生成の際の形態素解析には MeCab [⋆3]を使用し，名詞および形容詞のみを特徴ベクトルの次元として用いた．また特徴ベクトルの重み付けには収集された関連 QA 対の質問文および回答文をコーパスとして算出した TF-IDF 値を用いた．

⋆1 Yahoo! 知恵袋．http://chiebukuro.yahoo.co.jp/
⋆2 Yahoo! 知恵袋・質問検索 API．http://developer.yahoo.co.jp/webapi/chiebukuro/
⋆3 MeCab．http://mecab.sourceforge.net/

表 1　QA 対の信憑性の評価に用いたテストセットの例．質問はファクト型に限定されている．各質問で下線の引かれた語は関連 QA 対の検索に使用された語．#T，#F，#N は各質問への回答のうち著者が実際に正しい（true），誤っている（false），どちらか判定不能（neutral）と判定した回答の個数

Table 1　Examples in test set for evaluating QA pairs' credibility. The type of the questions in the test set is factoid. The underlined terms in each question were used to search for related QA pairs. #T, #F, and #N means the numbers of answers that we judged to be, true, false, and neutral (difficult to judge as true or false), respectively.

質問（一部抜粋）	全回答数（#T/#F/#N）
なぜ ナメクジ に 塩 をかけると死んでしまうのですか？	11 (4-5-2)
動いてる 電車 の中で ジャンプ しても，その場に着地できるのはなぜですか？なぜジャンプ中に電車だけ進まないのですか？	6 (3-2-1)
携帯電話 の 電波 が本当に ペースメーカー に影響を与えるのですか？	9 (4-4-1)
薬って，何で 水 で 飲ま ないといけないんですか？胃に入れば一緒だから，水分ならなんでもいいと思うんですけどね...	8 (4-1-3)
化学 でいう「イオン」って 何 なんですか？	3 (1-1-1)
重力 と 引力 の 違い を簡単に説明したいのですが？	2 (1-1-0)
豆腐 を 数える 時はなんで，一 丁 二丁なんです???	2 (1-1-0)
米酢 と 穀物酢 の 違いってなんですか？また，レシピにある「酢」とはどちらのことをいうのでしょう？	1 (1-0-0)
洋食器 の マイセン は どこ のブランド？	5 (5-0-0)
馴染みの歯医者さんで 歯石 をとるだけだと 保険 は 適用外 になるのでしょうか	3 (1-2-0)

図 6 に，関連 QA 対集合を絞り込む際に使用した類似度閾値ごとの QA 対の信憑性スコアの分布と QA 対の回答の真偽との関係を記す．グラフの横軸は信憑性スコアを表し，縦軸はある信憑性スコアを持つと評価された QA 対のうち，実際に回答が正しかった QA 対，誤っていた QA 対，どちらとも判断できなかった QA 対の個数を表している．図 6 の各グラフを見ると，スコアの分布傾向に若干の違いが見てとれるが，すべてのグラフにおいて，QA 対の信憑性スコアが高ければ高いほど，実際に回答が正しい QA 対の割合が大きくなる傾向があった．また，信憑性スコアが小さいほど，実際に回答が誤っている QA 対の割合が大きくなる傾向があった．このことから，信憑性スコアと QA 対の回答の真偽との間に関係があることが予想される．そこで，「回答が正しい QA 対と回答が誤っている QA 対の信憑性スコアの平均に差がない」という帰無仮説を立てて t 検定を行った．2 つのスコアの分散に差がある場合はウェルチの t 検定，分散に差がない場合スチューデントの t 検定を行った．分散に差があるか否かを調べるには F 検定を行った．結果，有意水準 0.05 で帰無仮説は棄却され「回答が正しい QA 対と回答が誤っている QA 対の信憑性スコアの分布には有意な差がある」ことが統計的に確認された．この結果から，ある閾値 t 以上の信憑性スコアを持つ QA 対は回答が正しい QA 対である，あるいは閾値 f 以下の信憑性スコアを持つ QA 対は回答が誤っている QA 対であると設定することで，ある程度の精度でシステムに回答の真偽を自動判定できる可能性がある．

そこで，QA 対の信憑性スコアを用いた回答の真偽の自動判定能力に関して評価を行った．評価指標としては適合率（precision），再現率（recall），F 値を用いた．評価に用いた $precision_{true}(t)$，$recall_{true}(t)$，$precision_{false}(f)$，$recall_{false}(f)$ は以下のように定義した：

$$precision_{true}(t) = \frac{\sharp(True_{real} \cap True_{system(t)})}{\sharp True_{system(t)}} \tag{11}$$

$$recall_{true}(t) = \frac{\sharp(True_{real} \cap True_{system(t)})}{\sharp True_{real}} \tag{12}$$

$$precision_{false}(f) = \frac{\sharp(False_{real} \cap False_{system(f)})}{\sharp False_{system(f)}} \tag{13}$$

$$recall_{false}(f) = \frac{\sharp(False_{real} \cap False_{system(f)})}{\sharp False_{real}} \tag{14}$$

ここで，$\sharp(X)$ は集合 X の要素数を，$True_{real}$ は実際に回答が正しい QA 対の集合を，$True_{system}(t)$ は閾値 t 以上の信憑性スコアを持つ QA 対の集合を表している．同様に $False_{real}$ は実際には回答が誤っている QA 対の集合を，$False_{system}(f)$ は閾値 f 以下の信憑性スコアを持つ QA 対の集合を表している．閾値 t，f は真偽判定方法の性質上 $t \geq f$ を満たす．

関連 QA 対の絞り込みに用いる類似度閾値ごとに信憑性スコアに対する閾値 t，f を変化させ，実際に回答の正しい QA 対および実際に回答が誤っている QA 対の検出性能の評価を補完適合率（interpolated precision）と再現率（recall）の観点から行った．ここで，補

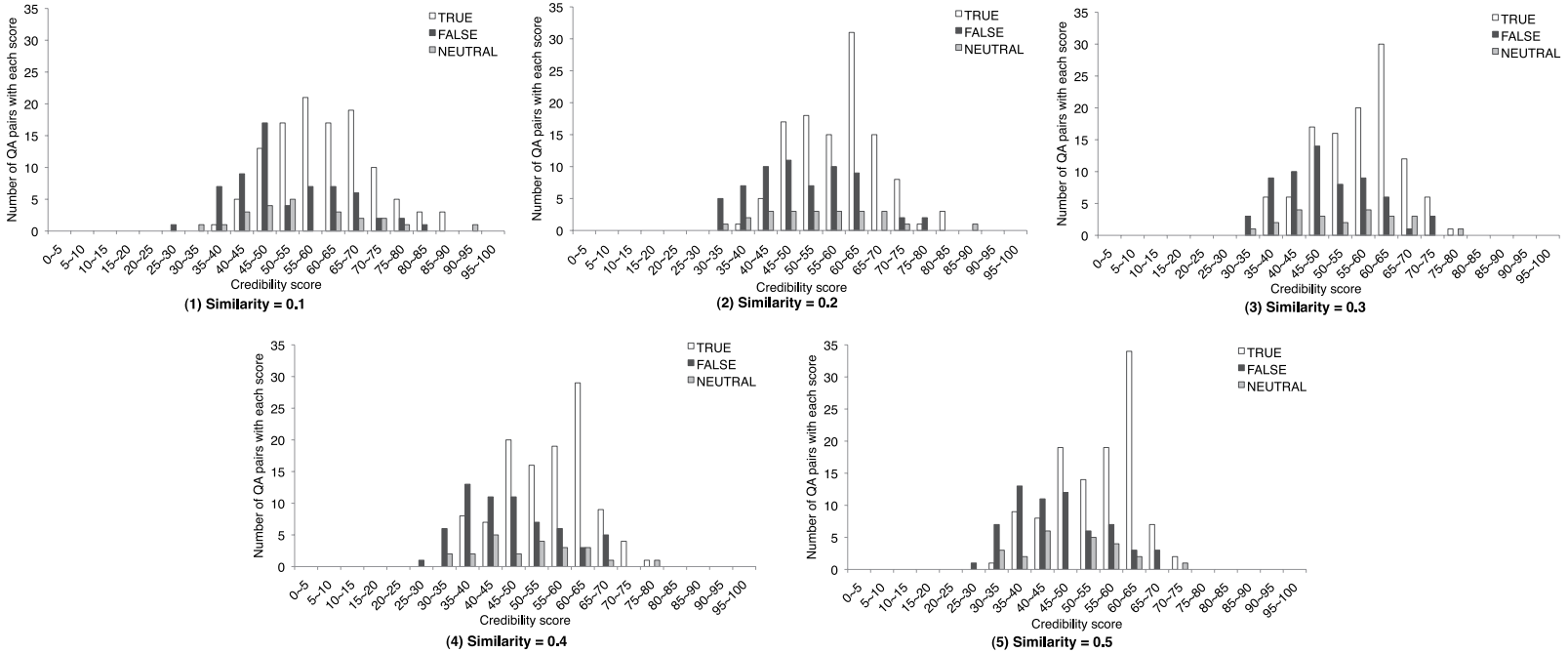
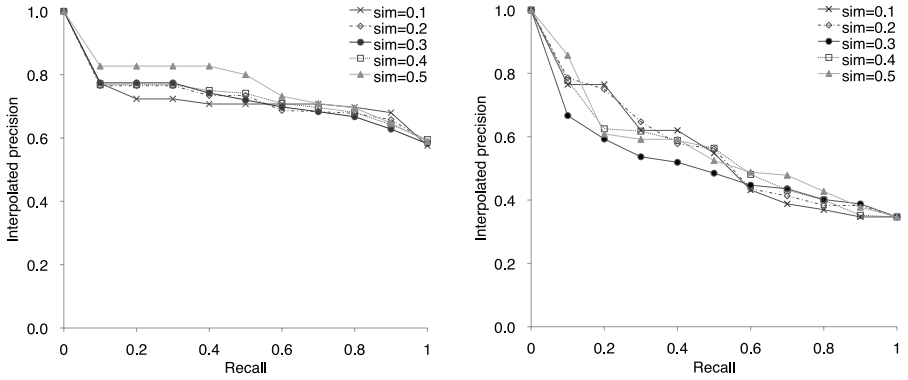


図 6　関連 QA 対の絞り込みに用いる類似度閾値ごとの信憑性スコアの分布
Fig. 6　Histogram of credibility scores in performing the credibility analysis.

完適合率 $precision_{interp}$ はある再現率 r における適合率を $precision(r)$ とした際，$r' \geq r$ において $precision_{interp}(r) = max_{r \geq r'} precision(r')$ と定義されるものである．この定義を用いて作成した 11 点補完適合率グラフを**図 7** に記す．図 7 のグラフ中の各点は，(a) のグラフではある信憑性スコア t 以上の QA 対の回答を正しいと判定したときの補完適合率と再現率の関係を，(b) のグラフではある信憑性スコア f 以下の QA 対の回答を誤っていると判定したときの補完適合率と再現率の関係を示している．グラフ (a) によると，回答が正しい QA 対の検出に関しては，補完適合率と再現率の両方の観点から評価した場合，類似度閾値を 0.5 に設定したケースが最も良い性能を示した．他の類似度閾値に関しては 0.1 に設定した場合が最も性能が悪かったが，0.2，0.3，0.4 に関してはほぼ同等の検出性能を示した．一方，グラフ (b) によると，実際に回答が誤っている QA 対の検出に関しては，適合率・再現率のバランス傾向は類似度閾値によって大きく異なり，最適な類似度閾値を決定することは難しい．システムによる真偽判定は，実用を考慮すると高い適合率が要求される．そこで，補完適合率が高い状況（$precision_{interp} \geq 0.7$）に着目すると，最適な類似度閾値は回答が正しい QA 対と回答が誤っている QA 対の両方の検出性能に関して適合率と再現率のバランス良く高い 0.2 もしくは 0.5 と考えられる．

関連 QA 対集合の絞り込みに用いる類似度閾値を下げた場合，ターゲットとなっている QA 対の質問に実際には関係のない QA 対が収集されてしまう可能性がある．そのようなケースでは，ターゲット QA 対の回答が実際に正しい場合でも，信憑性スコア計算の段階で無関係な QA 対から負の影響を受けてしまい，結果的に信憑性スコアが低くなってしまう．一方，閾値を上げすぎると，信憑性分析に必要な関連 QA 対が集まらず，ターゲット



(a) 実際に回答が正しい QA 対の検出性能　(b) 実際に回答が誤っている QA 対の検出性能

図 7　関連 QA 対の絞り込みに用いる質問類似度閾値ごとの 11 点補完適合率グラフ

Fig. 7　Interpolated precision graph to show the perfomance of QA pairs's credibility analysis with different question thresholds.

QA 対の信憑性スコアを適切に計算できなくなる．よって，関連 QA 対の絞り込みに用いる類似度閾値としては中間程度のものが望ましい．また，現実問題として，「正しい回答を誤っていると判断してしまうこと」よりも「誤っている回答を正しいと判断してしまうこと」の方がリスクが大きい．信憑性分析に必要な関連 QA 対の収集の際には，回答が正しい QA 対の検出性能よりも回答が誤っている QA 対の検出性能を重視するべきである．それゆえ，図 7 の結果を考慮すると，今回の実験に関しては関連 QA 対の絞り込みに用いる類似度閾値としては適合率と再現率がバランス良く高かった 0.2 に設定するのが最適であるといえる．類似度閾値を 0.2 に設定した際に適合率が 0.7 以上で F 値が最大となる信憑性スコア閾値 t，f について調べた．実際に正しい回答を正しいと自動判定するための信憑性スコア閾値 t は 60，その設定における適合率は 0.734，再現率は 0.509 であった．また，実際に誤った回答を誤っていると自動判定するための信憑性スコア閾値 f は 41，その設定における適合率は 0.700，再現率は 0.222 であった．

最後に，QA 対の信憑性スコアを用いた回答の真偽の自動判定に失敗したケースについて述べる．実際に回答が正しい QA 対を「回答が誤っている」と誤判定されてしまった QA 対を確認すると，意味的には正しいことを述べているが，説明に用いられている語彙が他の正しい回答とは異なっている回答が散見された．たとえば，「水は 100 度で沸騰するが 40 度程度の風呂でも湯気が発生する理由」を問う質問に対しては「湯気は水蒸気でなく液体である」ことを示唆する説明が回答に期待される．しかし，QA サイト内では同様の理由を説明するために「水蒸気」という言葉を用いている回答もあれば単に「気体」という言葉を用いたり，「液体」を説明するのに「水の粒」と説明している回答があるといったように，回答によって表現が様々であった．このため特徴ベクトル間の類似度では内容の類似度をうまく測ることできず，結果，全体的に QA 対の信憑性スコアが低くなってしまい回答の真偽の誤判定が起こってしまったと考えられる．実際に誤っているにもかかわらず「正しい」と誤判定されてしまった回答を確認したところ，結論を示すために用いられる語は異なるが，回答内容の大部分は正しい回答で用いられている語と共通していた．このため，実際には誤った回答であるにもかかわらず，回答が正しい QA 対から受ける support 値が大きくなり，結果的に回答が誤っている QA 対が大きい信憑性スコアを得てしまい，誤判定が起こってしまったと考えられる．また，そもそも QA サイト内に類似質問が少ない質問を持つ QA 対の真偽判定を行った場合，真偽判定に失敗するケースが散見された．これは，類似質問が少ない場合，分析に必要な関連 QA 対の量が十分でないために QA 対間の support 関係分析が十分に行えないため妥当性の低い信憑性スコアが算出されてしまったと予想される．

4. 画像・テキスト対に着目した画像の信憑性

「百聞は一見にしかず」という諺にもあるように，画像コンテンツはテキストコンテンツの内容理解を助けるのに非常に効果的である．しかし，画像がねつ造されている，使用されている画像が古い，画像とテキスト内容の間に整合性がない，といったように，画像によって語られる内容に誤りがある場合は読み手に多大な影響を与えかねない．

本章では，テキストと画像の対に着目した画像信憑性分析について述べる．本章では，2 種類の画像・テキスト対の信憑性を取り上げ，2 章で提案した信憑性分析モデルを画像・テキスト対の信憑性分析に適用するための具体的な実装について述べる．

4.1 テキストに対する画像の典型性

画像の信憑性に着目した際，最も深刻なケースの 1 つはラベルやキャプションといったテキスト内容に対して説明のために使用されている画像が適切でないケースである．たとえばオークションサイトなどでは，ある商品に関して実際の商品の様相とは異なる画像が使用されていることがしばしば見受けられる．

このような画像を検出する方法として，ターゲットとなっている商品に関する画像を収集し，ターゲット画像と収集された画像とを比較することが考えられる．ターゲット画像が商品を表す画像として正しい場合，収集された画像中に類似する画像が多数発見されるべきで

ある．このような確認法は，ある商品に関するターゲット画像の典型性を評価していると考えることができる．このことから，商品名のようなエンティティ名に対する画像信憑性の評価に関して以下のような仮説を設ける：

- あるエンティティ名に関する画像が与えられたとき，ターゲット画像がそれと類似しているほど，ターゲット画像はエンティティ名に対する画像として信憑性が高まる．
- そのような画像が多数存在するほど，ターゲット画像の信憑性はより高まる．

上記の仮説は画像・テキストの対に着目した画像の信憑性分析ととらえることができる．2 章で提案したモデルを用いて，上記の仮説に基づいた画像の信憑性分析を定式化する．画像・エンティティ名対 $p_1 = (e_1, i_1)$，$p_2 = (e_2, i_2)$ が与えられたとき，エンティティ名 e_1，e_2 間の距離 $dist_1(e_1, e_2)$ および画像 i_1，i_2 間の距離 $dist_2(i_1, i_2)$ を以下のように定義する：

$$dist_1(e_1, e_2) = \begin{cases} 0, & \text{if } e_1 = e_2 \\ 1, & \text{otherwise} \end{cases} \tag{15}$$

$$dist_2(i_1, i_2) = 1 - sim_v(i_1, i_2) \tag{16}$$

ここで $sim_v(i_1, i_2)$ は画像 i_1，i_2 間の画像類似度を表す．今回，仮説を立てた時点では，テキストの種類として “エンティティ名” を想定した．よって，画像・テキスト対に着目した画像の信憑性分析に必要な関連データ対集合としては，ターゲットのエンティティ名と同じエンティティ名を持つ画像・エンティティ名対のみが対象となる．この場合，support 値 $sup(p_1, p_2)$ は式 (3) を用いて以下のように再定義される：

$$sup(p_1, p_2) = sup_{dom}(p_1, p_2) = sim_v(i_1, i_2) \tag{17}$$

最終的に，式 (2)，(6)，(17) をターゲット画像・エンティティ名対およびその関連画像・エンティティ名対集合に適用することで計算することで，画像・エンティティ名対の信憑性スコアが計算される．分析に必要となる画像・エンティティ名対を収集するには，ターゲットとなっているエンティティ名をクエリとして画像データベースに問合せをすることで対応できる．このような手順によってエンティティ名に対する画像の典型性を評価する操作は，Jing らによって提案された VisualRank アルゴリズム[18] に相当すると考えられる．

4.2　テキストに対する画像の整合性

テキストに対する画像の典型性は画像・テキスト対に着目した画像の信憑性を評価するうえで有用な評価観点の 1 つではあるが，それだけでは画像の信憑性を厳密に評価できないケースも存在する．例として，**図 8** のように類似する画像に対して別人物の名前がラベルとして用いられているケースがあげられる．この例では，ウェブページ A では画像上の人物が「足利尊氏」，ウェブページ B では画像上の人物が「高師直」と言及されている．実際に「足利尊氏」というクエリでウェブ画像検索を行った場合，検索結果の上位の大半にウェブページ A の画像と類似する画像が含まれている．それゆえ，画像の典型性をテキストに対する画像の信憑性の評価基準として用いた場合，ウェブページ A 中の画像は「足利尊氏」の画像として信憑性が高いということになる．しかし，「高師直」という人物名でウェブ画像検索を行った場合も図 8 の画像と類似する画像が多数得られてしまう．実際，近年の日本史研究では，図 8 の画像の人物は「足利尊氏」ではなく「高師直」の可能性が高いということが指摘されている．この事例からも分かるように，より厳密に画像の信憑性を評価するためには，エンティティ名に対する画像の典型性を評価するだけでは不十分であり，別エンティティ名に関する画像を収集した際に，ターゲット画像と類似する画像上で異なるラベルが衝突していないか否かを確認する必要がある．この種の評価は，画像・テキスト間の整合性に着目した画像・テキスト対の信憑性分析ととらえることができる．以下では，2 章で提案したモデル上でテキストに対する画像の整合性を評価する方法について述べる．さらに，画像・テキスト対の一例として画像と歴史人物名の対を取り上げ，その信憑性を整合性の観点から評価するための現実的な実装についても述べる．


ページ A

足利尊氏は朝廷から征夷大将軍として任命され
室町幕府を創設した．


ページ B

肖像画の人物は足利尊氏ではなく，尊氏の付き人で
であった **高師直**である．

図 8　画像とそのキャプションに書かれた内容が矛盾している例
Fig. 8　Case where images and their captions are not consistent.

4.2.1　アプローチ

ここでは，画像・テキスト対として画像・エンティティ名対を取り上げ，画像の信憑性を画像・エンティティ名間の整合性の観点から評価することを考える．ターゲット対のエンティティ名に対応する画像を収集した際，多数の画像がターゲット対の画像と類似し，同時にターゲット対の画像に類似する画像に別名のエンティティ名が対応付けされているような状況がなければ，ターゲット対は整合性が高いと考えることができる．このことから，エンティティ名に対する画像の信憑性分析に関する以下のような仮説を設ける：

- ターゲット対のエンティティ名に関する画像を収集したとき，多数の画像がターゲット対の画像と類似していれば，ターゲット対の信憑性は高まる．
- ターゲット対のエンティティ名とは異なるエンティティ名に関する画像を収集したとき，多数の画像がターゲット対の画像と類似していなければ，ターゲット対の信憑性は高まる．

この仮説に従った信憑性評価は，2 章で提案した **dominance モデル**と **uniqueness モデル**の両方を考慮した信憑性評価として定式化することができる．その際，support 値 $sup(p_1, p_2)$ は式 (5)，(15)，(16) を用いて以下のように定義される：

$$sup(p_1, p_2) = \alpha \cdot sup_{dom}(p_1, p_2) + \beta \cdot sup_{uni}(p_1, p_2) \quad (18)$$

$$= \alpha(1 - dist_1(e_1, e_2)) \cdot sim_v(i_1, i_2) + \beta \cdot dist_1(e_1, e_2) \cdot (1 - sim_v(i_1, i_2)) \quad (19)$$

ターゲットである画像・エンティティ名対の信憑性スコアは式 (2)，(6)，(19) をターゲット対およびその関連画像・エンティティ名対集合に適用することで評価することができる．パラメータ α および β は $\alpha + \beta = 1$ を満たす範囲で自由に調整することができる．

4.2.2　現実的な実装

実際に画像の信憑性を画像・エンティティ名間の整合性の観点から評価するには，関連する画像・エンティティ名対の収集と画像類似度の計算が非常に重要な鍵となる．ここでは，画像・エンティティ名対の一例として図 8 で取り上げたような画像・歴史人物名対に焦点を当て，画像・エンティティ名間の整合性の観点から画像の信憑性分析を行う現実的な実装方法について述べる．以下，システムのフローを示す：

(1)　信憑性分析のターゲットとなる画像・歴史人物名対を入力として与える．
(2)　入力されたターゲット対の歴史人物名をクエリとしてウェブ画像検索を行い，同歴史人物名に関する画像を収集する．
(3)　事前に用意された歴史人物名リストを用いて，収集された画像の周辺テキストから別の歴史人物名を抽出する．
(4)　収集された歴史人物名に関する画像をステップ (2) と同じ方法で収集する．
(5)　収集された画像およびターゲット対の画像から類似度計算のための画像特徴量を抽出する．
(6)　収集された画像・歴史人物名対とターゲット対に式 (2)，(6)，(19) を適用し，ターゲット対の信憑性スコアを計算する．

信憑性分析に必要な関連データ対の収集法と画像類似度の計算手法について以下に記す：

画像特徴量の抽出と類似度計算

画像から画像特徴量を抽出する手法は多数提案されている．近年では，色特徴量やテクスチャ特徴量のような大域的特徴量よりも画像のスケールや回転に無関係な特徴をとらえることができる局所特徴量が注目されている．今回の提案手法では，近年最も注目されている SIFT（scale invariant feature transform）特徴量[23] を用いて画像特徴量を表現し，画像間の類似度を計算する．画像に SIFT アルゴリズムを適用すると，キーポイントと呼ばれる画像の局所特徴量をよく表現する特徴点が複数抽出される．キーポイントは 128 次元のベクトルで記述されている．提案する信憑性分析手法では，画像 i_1 と i_2 が共有するキーポイントの数を画像 i_1 と i_2 が持つキーポイントの総数の平均で割った値を画像 i_1 と i_2 の間の類似度として定義する．

画像・歴史人物名対の収集

ある画像に対応付けされた歴史人物名が誤りである場合，ウェブ上に存在するその類似画像の周辺テキストに別人物の名前が出現することが期待される．そこで，別人物に関する画像を収集するために，ターゲット対の歴史人物名でウェブ画像検索を行い，その周辺テキストから別の歴史人物名を抽出する．その際，事前に用意した 7225 名が掲載された歴史人物名リストを用いて，そこに登場する人物のみを別人物候補として抽出する．人物名を収集後，収集した各人物名をクエリとしてウェブ画像検索を行い，上位 N 件を取得する．最終的に，画像検索に用いたクエリと収集された画像のペアをターゲット対の関連画像・歴史人物名対とする．

現実問題としてウェブページである歴史人物に関する記述が行われる場合，表記揺れ（豊臣秀吉 → 豊富秀吉）や人物名の短縮（豊臣秀吉 → 秀吉），同じ人物を意味するが別の呼称が使用されるケース（聖徳太子 → 厩戸皇子）などが考えられる．本アプローチでは単純化のためにこの種の問題は扱わず，あるエンティティどうしの表記が異なれば別人物として扱う．提案モデルではエンティティ間の距離は柔軟に定義可能であるので，Oyama らが提案するオブジェクト識別手法[30]，Ahmad らの表記揺れを考慮したクエリ間距離の計算方法[3] などを用いることでエンティティ間の距離をより詳細に定義することが可能である．

4.2.3　評　　価

画像とエンティティ名間の整合性に着目した画像の信憑性分析手法を評価するために実データを用いて評価実験を行った．実験の目的は算出された画像・エンティティ名対の信憑性スコアが画像の真偽判定にどの程度有効であるか評価することである．

歴史人物画像とそれに対応付けられたラベル間の正誤フラグが付与されたテストセットは我々の知る限り存在しない．そこで，テストセットとしては，歴史上の人物に関する画像のうち実際に真偽が明らかなものを歴史書籍[⋆1]を参考に選んだ．最終的にテストセットとして用意された画像・エンティティ名対は 35 個で，そのうち 22 個が実際に正しい画像，13 個が実際は誤った画像が使用されていた画像・エンティティ名対であった．

実験では，歴史人物の画像を収集する際に Google Ajax Search API [⋆2]を用いた．1 クエリあたりの画像検索結果の取得件数は最大 30 個に設定した．画像間類似度計算には速度を重視するために，また確実に検索結果の画像を取得するために，オリジナル画像ではなく API が返すサムネイル画像を用いた．式 (19) 中の α および β はそれぞれ 0.5 に設定した．このような条件下で，テストセット中の各画像・エンティティ名対の信憑性スコアを計算した．実験では，式 (19) で $\alpha = 1$，$\beta = 0$ に設定し dominance のみを考慮した信憑性スコア（4.1 節で述べたエンティティ名に対する画像典型度）および式 (19) で $\alpha = 0$，$\beta = 1$ に設定し uniqueness のみを考慮した信憑性スコアの計算も行い，整合性に着目した信憑性分析のスコアとの比較を行った．

図 9 に整合性に注目した (a) 画像・エンティティ名対の信憑性スコアの分布，(b) 画像典型性に着目した画像・エンティティ名対の信憑性スコアの分布，(c) 画像・エンティティ名対の uniqueness に着目した信憑性スコアを示す．グラフによって実際に画像が正しく使用されていた画像・エンティティ名対のスコアの分布と画像が誤って使用されていた画像・エンティティ名対のスコアの分布を確認することができる．図 9 の (a) によると，画像・エンティティ名対の dominance と uniqueness の両方を考慮した整合性に着目した信憑性スコアは画像の真偽と関連があることが予想される．すなわち，画像・エンティティ名対の信憑性スコアが大きいほど，実際に画像が正しく使用されている画像・エンティティ名対の割合が大きくなり，信憑性スコアが小さいほど，実際に画像が誤って使用されている画像・エンティティ名対の割合が大きくなる傾向が確認できた．画像が正しく使用されている画像・エンティティ対と画像が誤って使用されている画像・エンティティ名対の信憑性スコアの分布に対して t 検定を行ったところ，統計的な有意差が確認された（$p = 4.79 \times 10^{-5} < 0.05$）．この結果から，3.2 節で述べた実験と同様信憑性スコアに対してある閾値を設定することで，画像の真偽をある程度の精度で自動判定できる可能性がある．一方，図 9 の (b) によると，

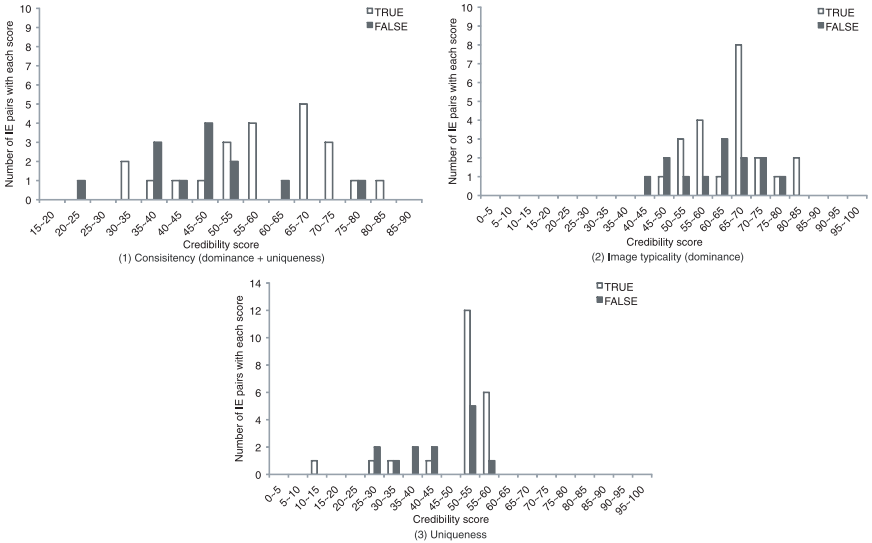


図 9　各手法に応じた画像・エンティティ名対の信憑性スコア分布
Fig. 9　Histogram of credibility scores of image-entity name pairs in using different methods.

エンティティ名に対する画像典型性を考慮した信憑性スコア（dominance のみを考慮したスコア）を算出した場合，実際に画像が正しく使用されている画像・エンティティ名対と実際に画像が誤って使用されている画像・エンティティ名対の信憑性スコアの分布の間に統計的な有意差は確認できなかった（$p = 1.61 \times 10^{-2} > 0.05$）．信憑性スコアが高かったが実際の画像が誤っている画像・テキスト対を確認したところ，大半は類似画像に別人物の名前が関連づけられているものであった．また，図 9 の (c) によると，画像・エンティティ名対の uniqueness のみを考慮した信憑性スコアに関しても，実際に画像が正しく使用されている画像・エンティティ名対と実際に画像が誤って使用されている画像・エンティティ名対の信憑性スコアの分布の間に統計的な有意差は確認できなかった（$p = 1.22 \times 10^{-1} > 0.05$）．信憑性スコアが高かったが画像・テキスト対の真偽判定に失敗したものを確認したところ，対象エンティティ名に無関係な画像が大半であった．Uniqueness のみを考慮した画像・エンティティ名の信憑性分析の場合，別人物の画像群に類似画像がなければスコアが高まる．そのため，本来ならばエンティティ名に対する画像として無関係な画像でも典型性が考慮されていないため不当に信憑性スコアが高くなってしまう．これらの結果から，厳密に画像の

⋆1 なぜ偉人たちは教科書から消えたのか．http://www.amazon.co.jp/なぜ偉人たちは教科書から消えたのか-【肖像画】が語る通説破りの日本史-河合-敦/dp/433497502X
⋆2 Google Ajax Search API. http://code.google.com/intl/ja/apis/ajaxsearch/

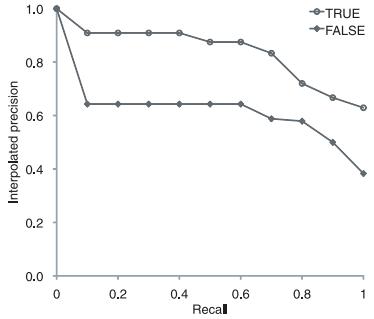


図 10　画像・エンティティ対の真偽検出の性能を示す 11 点補完適合率グラフ
Fig. 10　Interpolated precision graph to show the performance of image-entity name credibility analysis.

真偽を判定するためには dominance と uniqueness の両方を考慮して信憑性分析を行う必要があると考えられる．

整合性に着目した画像・エンティティ対の信憑性スコアを用いた画像の真偽の自動判定能力に関して，信憑性スコアに対する閾値 t，f を変化させ，エンティティ名に対して選択されている画像が実際に正しい画像・エンティティ名対およびエンティティ名に対して選択されている画像が実際に誤っている画像・エンティティ名対の検出性能の評価を補完適合率（interpolated precision）と再現率（recall）の観点から行った．評価には 3.2 節の実験で用いた式 (11)，(12)，(13)，(14) を用いた．ここで，$True_{real}$ は実際には画像が正しく使用されている画像・エンティティ名対の集合，$True_{system}(t)$ は信憑性スコアが閾値 t 以上であった画像・エンティティ名対の集合，$False_{real}$ は実際には画像が誤って使用されている画像・エンティティ名対の集合，$False_{system}(f)$ は信憑性スコアが閾値 f 以下であった画像・エンティティ名対の集合を表す．自動判定結果の 11 点補完適合率グラフを**図 10** に記す．図 10 のグラフ中の “TRUE” の線上の各点はある信憑性スコア t 以上の画像・エンティティ対の画像がエンティティ名に対して正しいと自動判定したときの補完適合率と再現率の関係，“FALSE” の線上の各点はある信憑性スコア f 以下の画像・エンティティ対の画像がエンティティ名に対して誤っていると自動判定したときの補完適合率と再現率の関係を示している．グラフによると，エンティティ名に対して正しい画像が選択されている画像・エンティティ名対を自動検出する性能は良好であった．適合率が 0.7 以上でかつ F 値が最大になる点を調べたところ，信憑性スコア t を 52 に設定したとき適合率 0.833，再現率が 0.636 となった．一方，エンティティ名に対して誤っている画像が選択されている画像・エンティティ名対を自動検出する精度は正しい画像が選択されている画像・エンティティ名対を自動検出する性能ほどは良好ではなかったものの，再現率が 0.6 以下の条件下では適合率 0.65 程度で誤った画像が選択されている画像・エンティティ名対を自動検出することができた．今回の実験で用いたテストセットは小規模であったため，画像の真偽判定に最適な信憑性スコアの閾値や自動判定能力について深く言及することはできないが，厳しい閾値を設けることで実用的な画像の真偽判定を期待できる．

エンティティ名に対して誤った画像が選択されている画像・エンティティ名対の検出に失敗したものを分析したところ，別人物の名前が周辺テキストから抽出されなかったケースが散見された．提案実装では画像に映し出された人物として別人物が考えられるならば周辺テキストに別人物の名前が出現するという仮説を置いていたが，うまく別人物の名前が抽出されなかった場合，uniqueness 度合いを評価する support 値が信憑性スコアの計算に反映されない．よって信憑性スコアの計算に失敗し真偽の判定に失敗したと考えられる．別人物の名前抽出に失敗したケースとしては大きく分けて 2 つのケースが確認された．1 つは別のウェブページではターゲットの人物名とは異なる別人物の名前をあげて類似画像を扱っているにもかかわらず収集した画像の周辺テキストから別人物名が抽出できなかったケースである．このようなケースに対応するためには別人物の収集方法を再検討する必要がある．もう 1 つは，画像に写ってる人物がターゲットの人物ではないと指摘されるにとどまり，代替人物の名前が明記されていないケースである．このようなケースの一例として「西郷隆盛」の肖像画があげられる．西郷隆盛の肖像画として最も有名な画像[⋆1]は，歴史研究者の間では，西郷隆盛を描いたものではない，ということが近年の主要な説となっている．このような説はウェブには浸透していないため，類似する画像に関して別人物名をあげ異論を唱えるウェブページはほとんど存在していない．それゆえ，類似画像の周辺テキストから別人物の名前を抽出することができず，本来低くなるべき信憑性スコアが高くなってしまった．提案手法はウェブ上の多数派情報を評価する手法であるため，このような状況には対応できない．

5. 考　　察

質問・回答対および画像・エンティティ名対を対象にした評価実験から，適切に条件を整えたうえで提案した信憑性分析モデルを用いることで，データ対の形で表現されるウェブ情

⋆1 西郷隆盛の肖像画．http://www.ndl.go.jp/portrait/datas/85_1.html

報の真偽をある程度の精度で判定することができることが示された．提案モデルを用いるためには，以下の 2 種類の条件を整えなければならなかった：(1) ウェブ情報の種類とそれが出現する状況に応じた信憑性評価観点（dominance と uniqueness のいずれか，またはその組合せ），(2) 分析ターゲットになっているデータと関連するデータ集合．

5.1 課　　題

信憑性評価観点に関してはより多様な信憑性分析のための support 関数の定義方法が検討課題としてあげられる．

本稿では，特定の観点からウェブ情報を信憑性を分析するためにデータ対間の support 関係として dominance，uniqueness を分析するための 2 種類の support 関係を導入したが，データ対間の距離に着目することでさらにもう 1 種類の support 関係を考えることができる．その例を**図 11** に示す．この例では，ある報道局によって書かれた記事の信憑性に着目しており，信憑性観点の 1 つとしてターゲット記事に類似する記事が世界中の様々な報道機関からも発行されているか否かを考えている．このケースでは記事の信憑性を分析するために記事・報道局のデータ対に着目することができ，2.1 節と同様にデータノード間の距離関係に着目すると，記事データ間の類似度が大きく，報道局ノード間の距離（たとえば報道局間の空間的距離）が大きい場合，データ対は他のデータ対から大きな support を得られると考えられる．このようなタイプの support 関係はデータ対の多様性（diversity）が信憑性観点として重要なケースを分析するのに有用である．2.1 節で述べた dominance，uniqueness と同様に，データノード間の距離を考慮して diversity 分析のための support 関係を定義することで，式 (1)，(2)，(6) を用いて提案モデル上で信憑性分析を行うことができる．また diveristy を dominance や uniqueness と組み合わせて信憑性分析を行いたい場合，式 (5) に diversity を考慮するための support 関数を組み込むことで対応することが可能である．このように Diversity 分析のための support 関係を導入することでより多様な信憑性分析が可能となる．

柔軟かつ正確な信憑性分析を行うためには信憑性観点の決定方法も重要な検討課題となる．本稿では，例題として取り上げた質問・回答対および画像・エンティティ名対に対して信憑性分析を行う際，信憑性分析の観点およびパラメータを手動で設定した．実際にユーザが提案モデルを用いて信憑性分析を行うためには，信憑性分析観点に基づいたデータ対を設定したうえでデータ対間の support 関係の設定が必要となる．データ対の設定に関しては対象となっているウェブ情報と同種のデータ（例：画像信憑性であれば画像）を片側のノードに，信憑性分析観点に関連するデータ（例：画像の信憑性をラベルに対する画像の典型性

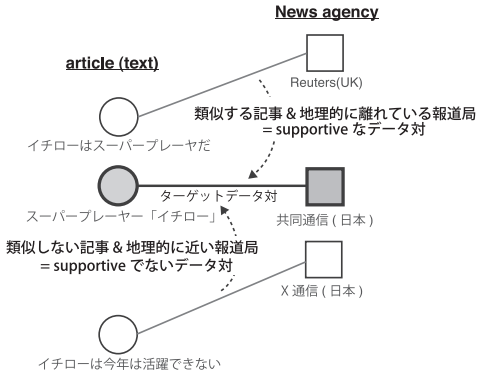


図 11　Diversity が信憑性分析の観点として重要になるケース
Fig. 11　Case where diversity is important as credibility aspect.

の観点から分析する場合はラベル）をもう片側のデータに設定すればよい．support 関係に関しては分析観点の特性とデータ対の距離関係を対応付けながら設定すればよい．しかし，ユーザがウェブ情報の信憑性を意識していないような状況では，信憑性分析を行うために必要なデータ対や分析観点，パラメータを決定すること自体がそもそも難しい．それゆえ，実用的にはウェブ情報の種類に応じて提案モデル上で信憑性分析を行うための最適な条件が自動的に推定されることが望ましい．自動的に信憑性観点やパラメータを決定する場合，対象としているウェブ情報の信憑性がユーザに依存しない「客観的な信憑性」なのか，もしくはユーザによって異なる「主観的な信憑性」であるかを考慮することも重要である[12]．今回実験で扱った信憑性は科学的な内容，事実といったように “credibility = correctness” と考えるウェブ情報の信憑性であった．そのようなタイプの信憑性の場合は何らかの方法で信憑性分析の観点およびパラメータを一意に決定すればよい．一方で，意見や主張といったようなユーザの価値観によって異なるようなタイプの信憑性を扱う場合，信憑性観点およびパラメータはユーザごとに異なるため一意に決めることはできない．このようなタイプの信憑性を分析する場合，ユーザからのフィードバック情報を基にユーザが重要と考える信憑性分析観点をユーザごとに推定し，それに応じて信憑性分析を行う必要がある．

関連データの収集も着目する信憑性観点に依存し，収集した関連データの質によって信憑性分析の精度も大きく左右される．しかし，実際は信憑性分析に必要なデータを適切に集めることは難しい．4.2.2 項で述べた画像・歴史人物名対の収集がその顕著な例である．

4.2.2 項では，類似する画像に対して異なる人物名が関連づけられているケースを扱った．画像・歴史人物名間の整合性を評価するために必要なデータ対の収集は，ターゲット人物名に関する画像の周辺テキストから別人物名を抽出し，抽出した人物名をクエリとして再度ウェブ画像検索を行うという方法をとった．しかし，ターゲットとなっている画像に類似する画像に別人物が関連づけられているかを調べることが目的であるならば，直感的には画像をクエリとするコンテンツベースの画像検索を行うべきである．実用レベルのコンテンツベースの画像検索エンジンが存在するならば，ターゲット画像に類似する画像を収集し，それらの画像が指すエンティティ名を周辺テキストやメタデータなどから効率良く抽出することができる．画像・テキスト対の信憑性を提案モデル上で分析する場合は，このようなコンテンツベースの画像検索エンジンが最適である．画像・テキスト対以外のデータ対を解析対象とする場合は，別種類の検索エンジンが必要となる．たとえば，ある報道局が報道している内容がバイアスがかかった情報なのかを diversity の観点から提案モデルを用いて分析する場合，類似内容を報道している他の報道局を検索する必要がある．このように，提案モデルで適切に信憑性分析を行うためには分析に必要なデータ対を検索する技術が重要となる．さらに，信憑性分析のロバスト性を確保するためには，収集するデータの量も重要となる．ケースによっては，大量のデータ対の関係性を表す大規模な行列に対する演算を行う可能性もある．そのようなケースでは，Kashima らが提案しているような高速かつスケーラブルな行列計算アルゴリズム[19] が必要となる．

5.2　提案モデルの限界

最後に提案モデルの限界について述べる．1 つ目の限界は，たとえば Q&A サイトでは質問によっては複数の正しい回答が存在しうる，といったように信憑性が高いデータ対として複数の候補が想定されるケース，つまりデータ間の関係として 1 対多関係が許される状況には提案モデルは対応できない，という点である．これは提案モデルでは，集合内のデータ対間の関係性に基づく信憑性分析を行う際に，データ間の妥当な関係とはあくまで 1 対 1 関係であるとの暗黙の仮定を置いていることに起因している．

2 つ目の限界は，集合知アプローチによって誤った信憑性分析を行ってしまう可能性があるという点である．提案モデルは「より信憑性が高いデータ対から多くのサポートを集めるデータ対は信憑性が高い」という集合知的アプローチによって信憑性分析を行うモデルである．それゆえ信憑性分析に用いるデータ対の大半が誤っているようなケースでは提案モデルでは妥当な分析を行うことができない．たとえば，2.1 節の図 4 で述べた画像の特異性を分析するケースにおいて，分析に用いる関連データ対がハンバーガー A の画像として誤ってフライドポテトの画像を掲載していたケースを考える．このようなケースでは本来分析に用いるデータ対として除外すべきフライドポテト画像がターゲットであるハンバーガー H の画像と類似していないため，ターゲットの画像の信憑性を不当に高めてしまうということが起こりえてしまう．掲載ミスを起こしている画像が少ない場合は，dominance 型の support を考慮することによってまず典型的なハンバーガー A の画像を求めておき，その後掲載ミスの画像を除外した関連データ対を用いて改めて uniqueness 分析を行うことで，先に述べたようなケースに対応しつつ特異性分析を行うことも可能である．しかし，そもそも掲載ミスの画像が大半であるような状況では代表画像を先に求めておく方法もうまく機能せず，対象となっているハンバーガー H の画像の特異性が不当に高められることになる．このような問題に対応するためには集合知的アプローチではない別の信憑性分析手法が必要となる．

6. 関 連 研 究

6.1　信憑性の評価観点

近年，信憑性研究は注目されはじめている．Fogg らは実験心理学の立場から信憑性の整理を行っており，様々な信憑性を「仮定に基づく信憑性」「外見に基づく信憑性」「評判に基づく信憑性」「経験に基づく信憑性」の 4 つに分類している[14]．Fogg らが焦点を当てているのは主にコンピュータの信憑性であるが，提案されている信憑性に対する考え方は情報の信憑性を考えるうえでも有用である．Fogg らは，ユーザがウェブサイトのどのような要素がユーザが信憑性を認知するうえで影響力が大きいかを明らかにするために大規模な調査も行っている[13]．Nakamura らは大規模なアンケート調査によって，ウェブ検索エンジンが信用できるか否かを考えるうえでユーザが重要視している要素を明らかにしている[28]．このように信憑性の要素の整理や調査は行われているものの，具体的な信憑性評価手法やそれに基づくアプリケーションなどの提案を行っている研究は少ない．

セマンティックウェブの研究分野で議論される Trust もウェブの信頼性と関連がある．セマンティックウェブで語られる Trust は，(a) 推論規則と推論のための論理を組み合わせることで評価されるファクトの信憑性，(b) 署名や暗号化によって担保される情報ソースの信頼性，として議論されている[4]．(a) の信憑性に関しては，事前に構築されたオントロジ，推論規則と推論のための論理がベースになるため，信憑性推定の対象は非常に小さい範囲に限定される．(b) の情報ソースの信頼性に関しては，実際に署名や暗号などのメタデータが付与されたウェブページが少ないため，実用的に評価するのが難しいのが現状である．

6.2 Trustworthiness と Expertise

特定のウェブ情報に対して特定の観点から信憑性分析，判断支援を行うシステムがいくつか提案されている．Vuong らは Wikipedia 記事上で生じている論争を自動的に検出する手法[36] を，Kittur らは Wikipedia 記事中で矛盾があると予想される文の可視化手法を提案し[21]，その可視化情報がユーザの信憑性判断にどのような影響を与えるかの評価を行っている[20]．これら Wikipedia の信憑性に関する研究は，記事の編集履歴に着目しており，trustworthiness の一要素として Wikipedia 記事のバイアス度合いを評価しているといえる．

ハイパーリンクやソーシャルアノテーションのような社会的評価に着目した研究も trustworthiness を評価する研究と見なすことができる．Yanbe らはウェブページにアノテーションされたソーシャルブックマークの数を利用してウェブページの評価を行う手法を提案している[38]．Asadi らはハイパーリンクを張っているウェブページの地域特性を考慮することで，ある地域におけるウェブページの人気度を算出するモデルを提案している[5]．Ding らはリンク元ウェブページの物理的位置を考慮することで，ウェブページの広域的な支持度を評価する手法を提案している[10]．これらの研究は一種の trustworthiness 評価ではあるが，ウェブ情報の真偽や妥当性の評価を直接目指しているわけではない．

Expertise の軸からのウェブページの評価手法を提案している研究も存在する．Nakatani らは Wikipedia をコーパスとして用いて，あるトピックに関するウェブページ中の話題の網羅度および詳細度を評価する手法を提案している[29]．B. Zhang らはウェブ検索の結果を用いることで，情報の豊富さの観点からウェブページを評価する手法を提案している[39]．また，J. Zhang らはオンラインコミュニティにグラフ解析を行うことで，ユーザの専門性を評価する手法を提案している[40]．本稿で提案した手法はあくまで trustworthiness に着目しているため，expertise の観点からウェブ情報の信憑性を評価することはできない．Expertise に基づく信憑性分析が有効に働くケースも存在するため，ケースに応じて trustworthiness, expertise の軸を使い分けて信憑性評価を行う必要がある．

6.3 質問応答コンテンツの品質とマルチメディアコンテンツの信憑性

近年，Q&A サイトにおける質問回答コンテンツおよびユーザの評価に関する研究がさかんになってきている．研究の大半が「ベストアンサー」のような高品質な回答の発見もしくは高品質な回答を投稿するユーザの発見に焦点が当てられている．Agichtein らは Q&A サイト上でのユーザ間のインタラクションを 2 部グラフで表現し，そのグラフを用いてユーザの authority スコアおよび hub スコアを求める手法を提案している[2]．Suryanto らは回答者の専門性を考慮した回答の質の評価方法を提案している[35]．Bian らが提案する手法[6] では，回答の質を評価するために，質問回答コンテンツ，Q&A サイト上でのユーザ間のインタラクション，回答に対するコミュニティ上でのフィードバック情報を統合した評価方法が着目されている．これらの研究では，回答の質はユーザの満足度，分かりやすさ，役立ち度，正確さなどの複数の要因から構成されるものとして暗黙的にとらえられている．本稿では，質問・回答対の信憑性評価に明確に焦点を当てており，その点で既存研究と目的が異なる．

近年マルチメディアの研究分野では，画像や映像に対する意図的な編集や別ソースから複製された画像・映像の検出に関する研究が注目を集めている．これらはマルチメディアコンテンツの信憑性分析の一種と考えることができる．Gloe らは不正な画像加工を検知するために，撮影に用いられたデジタルカメラの特性分析や画像の再サンプリングに基づく手法を提案している[15]．Zhu らは映像から意図的な映像コピーを検知する手法を提案している[41]．画像や映像の信憑性が危ぶまれるケースとしては，マルチメディアコンテンツに対する意図的な編集以外にも，画像や映像とそのキャプションや周辺テキストとの間に整合性がないケースもしばしば存在する．そのようなケースでは，本稿で提案したような画像・テキスト間の整合性を分析する評価手法が必要となる．

7. おわりに

本稿では，任意のデータの対で表現されるウェブ情報とその関連ウェブ情報間の support 関係の強さを分析することでウェブ情報の信憑性を評価するモデルを提案した．提案モデルでは，信憑性の評価観点として dominance および uniqueness の 2 つの基本的な観点を取り上げ，それぞれの観点に沿った信憑性分析方法を support 値の概念を導入することで定式化した．提案モデルは，信憑性分析の対象であるウェブ情報の種類やコンテキストに応じてウェブ情報をデータ対で表現し，support 値の定義を行うことで様々な種類のウェブ情報の信憑性を評価することができる．また，複数の信憑性観点を同時に考慮する必要がある場合も，support 値の定義をうまく構成することでモデルを崩すことなく対応できる．この意味において提案モデルは汎用的である．

また本稿では，提案モデルによって評価された情報の信憑性が実際の真偽判定にどの程度有効であるかを評価するために，質問・回答対および画像・歴史人物名対を取り上げ評価実験を行った．評価実験の結果から，提案モデルによって算出された信憑性スコアを使用することで，質問応答サイトに投稿された回答の真偽を適合率 70%以上で，歴史人物名に対する画像の真偽を適合率 65%以上で自動判定することができることが示された．

今後の課題としては，ウェブ情報の種類やそれが出現するコンテキストに応じた適切な信

憑性分析観点の自動推定方法の検討があげられる．また expertise の観点からウェブ情報の信憑性を評価するモデルの検討も重要な課題であると考えている．

謝辞　本研究の一部は，グローバル COE 拠点形成プログラム「知識循環社会のための情報学教育研究拠点」（研究代表者：田中克己），文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」，計画研究「情報爆発に対応するコンテンツ融合と操作環境融合に関する研究」（研究代表者：田中克己，課題番号 1809041），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」（研究代表者：田中克己），科学研究費補助金特別研究員奨励費「「Web 情報の信頼性の評価」に関する研究」（研究代表者：山本祐輔，課題番号 09J01243）によるものです．ここに記して謝意を表します．

参 考 文 献

1) Adamic, L.A., Zhang, J., Bakshy, E. and Ackerman, M.S.: Knowledge sharing and yahoo answers: everyone knows something, *Proc. 17th international conference on World Wide Web* (*WWW 2008*), pp.665–674 (2008).
2) Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G.: Finding high-quality content in social media, *Proc. international conference on Web search and web data mining* (*WSDM 2008*), pp.183–194 (2008).
3) Ahmad, F. and Kondrak, G.: Learning a spelling error model from search query logs, *Proc. conference on Human Language Technology and Empirical Methods in Natural Language Processing* (*HLT/EMNLP 2005*), pp.955–962 (2005).
4) Artz, D. and Gil, Y.: A Survey of Trust in Computer Science and the Semantic Web, *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol.5, No.2, pp.58–71 (2007).
5) Asadi, S., Zhou, X. and Yang, G.: Using Local Popularity of Web Resources for Geo-Ranking of Search Engine Results, *World Wide Web Journal*, Vol.12, No.2, pp.149–170 (2009).
6) Bian, J., Liu, Y., Agichtein, E. and Zha, H.: Finding the right facts in the crowd: factoid question answering over social media, *Proc. 17th international conference on World Wide Web* (*WWW 2008*), pp.467–476 (2008).
7) Cho, J. and Roy, S.: Impact of search engines on page popularity, *Proc. 13th international conference on World Wide Web* (*WWW 2004*), pp.20–29 (2004).
8) Cho, J., Roy, S. and Adams, R.E.: Page quality: in search of an unbiased web ranking, *Proc. 2005 ACM SIGMOD international conference on Management of data* (*SIGMOD 2005*), pp.551–562 (2005).
9) Denning, P., Horning, J., Parnas, D. and Weinstein, L.: Wikipedia risks, *Comm. ACM*, Vol.48, No.12, p.152 (2005).
10) Ding, J., Gravano, L. and Shivakumar, N.: Computing Geographical Scopes of Web Resources, *Proc. 26th International Conference on Very Large Data Bases* (*VLDB 2000*), pp.545–556 (2000).
11) Erkan, G. and Radev, D.: LexRank: Graph-based lexical centrality as salience in text summarization, *Journal of Artificial Intelligence Research*, Vol.22, No.2004, pp.457–479 (2004).
12) Flanagin, A.J. and Metzger, M.J.: The credibility of volunteered geographic information, *GeoJournal*, Vol.72, No.3, pp.137–148 (2008).
13) Fogg, B.J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P. and Treinen, M.: What makes Web sites credible?: a report on a large quantitative study, *Proc. SIGCHI conference on Human factors in computing systems* (*CHI 2001*), pp.61–68 (2001).
14) Fogg, B.J. and Tseng, H.: The elements of computer credibility, *Proc. SIGCHI conference on Human factors in computing systems* (*CHI 1999*), pp.80–87 (1999).
15) Gloe, T., Kirchner, M., Winkler, A. and Böhme, R.: Can we trust digital image forensics?, *Proc. 15th international conference on Multimedia* (*MM2007*), pp.78–86 (2007).
16) Harper, F.M., Moy, D. and Konstan, J.A.: Facts or friends?: distinguishing informational and conversational questions in social Q&A sites, *Proc. 27th international conference on Human factors in computing systems* (*CHI 2009*), pp.759–768 (2009).
17) Haveliwala, T.H.: Topic-sensitive PageRank, *Proc. 11th international conference on World Wide Web* (*WWW 2002*), pp.517–526 (2002).
18) Jing, Y. and Baluja, S.: VisualRank: Applying PageRank to Large-Scale Image Search, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.30, No.11, pp.1877–1890 (2008).
19) Kashima, H., Oyama, S., Yamanishi, Y. and Tsuda, K.: On Pairwise Kernels: An Efficient Alternative and Generalization Analysis, *Proc. 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining* (*PAKDD 2009*), pp.1030–1037 (2009).
20) Kittur, A., Suh, B. and Chi, E.H.: Can you ever trust a wiki?: impacting perceived trustworthiness in wikipedia, *Proc. ACM 2008 conference on Computer supported cooperative work* (*CSCW 2008*), pp.477–480 (2008).
21) Kittur, A., Suh, B., Pendleton, B.A. and Chi, E.H.: He says, she says: conflict and coordination in Wikipedia, *Proc. SIGCHI conference on Human factors in computing systems* (*CHI 2007*), pp.453–462 (2007).
22) Liu, Y., Gao, B., Liu, T.-Y., Zhang, Y., Ma, Z., He, S. and Li, H.: BrowseRank:

letting web users vote for page importance, *Proc. 31st annual international ACM SIGIR conference on Research and development in information retrieval* (*SIGIR 2008*), pp.451–458 (2008).
23) Lowe, D.: Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, Vol.60, No.2, pp.91–110 (2004).
24) McKnight, D.H. and Kacmar, C.J.: Factors and effects of information credibility, *Proc. 9th international conference on Electronic commerce* (*ICEC 2007*), pp.423–432 (2007).
25) Meola, M.: Chucking the Checklist: A Contextual Approach to Teaching Undergraduates Web-Site Evaluation, *portal: Libraries and the Academy*, Vol.4, No.3, pp.331–344 (2004).
26) Metzger, M.J.: Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research, *Journal of the American Society for Information Science and Technology*, Vol.58, No.13, pp.2078–2091 (2007).
27) Metzger, M.J., Flanagin, A.J. and Zwarun, L.: College student web use, perceptions of information credibility and verification behavior, *Computer & Education*, Vol.41, No.3, pp.271–290 (2003).
28) Nakamura, S., Konishi, S., Jatowt, A., Ohshima, H., Kondo, H., Tezuka, T., Oyama, S. and Tanaka, K.: Trustworthiness Analysis of Web Search Results, *Proc. 11th European Conference on Research and Advanced Technology for Digital Libraries* (*ECDL 2007*), pp.38–49 (2007).
29) Nakatani, M., Jatowt, A., Ohshima, H. and Tanaka, K.: Quality Evaluation of Search Results by Typicality and Speciality of Terms Extracted from Wikipedia, *Proc. 14th International Conference on Database Systems for Advanced Applications* (*DASFAA 2009*), pp.570–584 (2009).
30) Oyama, S. and Tanaka, K.: Learning a Distance Metric for Object Identification Without Human Supervision, *Proc. 10th European Conference on Principles and Practice of Knowledge Discovery in Databases* (*PKDD 2006*), pp.609–616 (2006).
31) Rieh, S.Y.: Judgement of information quality and cognitive authority in the Web, *Journal of the American Society for Information Science and Technology*, Vol.53, No.2, pp.145–161 (2002).
32) Rieh, S.Y. and Hilligoss, B.: College Students' Credibility Judgments in the Information-Seeking Process, *Digital Media, Youth and Credibility*, MIT Press, pp.49–71 (2007).
33) Sillence, E., Briggs, P., Fishwick, L. and Harris, P.: Trust and mistrust of online health sites, *Proc. SIGCHI conference on Human factors in computing systems* (*CHI 2004*), pp.663–670 (2004).
34) Stiff, J. and Mongeau, P.: *Persuasive communication*, Guilford Publications (2002).
35) Suryanto, M.A., Lim, E.P., Sun, A. and Chiang, R.H.L.: Quality-aware collaborative question answering: methods and evaluation, *Proc. 2nd ACM International Conference on Web Search and Data Mining* (*WSDM 2009*), pp.142–151 (2009).
36) Vuong, B.-Q., Lim, E.-P., Sun, A., Le, M.-T. and Lauw, H.W.: On ranking controversies in wikipedia: models and evaluation, *Proc. international conference on Web search and web data mining* (*WSDM 2008*), pp.171–182 (2008).
37) Wathen, C. and Burkell, J.: Believe it or not: factors influencing credibility on the Web, *Journal of the American Society for Information Science Technology*, Vol.53, No.2, pp.134–144 (2002).
38) Yanbe, Y., Jatowt, A., Nakamura, S. and Tanaka, K.: Can social bookmarking enhance search in the web?, *Proc. 7th ACM/IEEE-CS joint conference on Digital libraries* (*JCDL 2007*), pp.107–116 (2007).
39) Zhang, B., Li, H., Liu, Y., Ji, L., Xi, W., Fan, W., Chen, Z. and Ma, W.-Y.: Improving web search results using affinity graph, *Proc. 28th annual international ACM SIGIR conference on Research and development in information retrieval* (*SIGIR 2005*), pp.504–511 (2005).
40) Zhang, J., Ackerman, M.S. and Adamic, L.: Expertise networks in online communities: structure and algorithms, *Proc. 16th international conference on World Wide Web* (*WWW 2007*), pp.221–230 (2007).
41) Zhu, J., Hoi, S.C., Lyu, M.R. and Yan, S.: Near-duplicate keyframe retrieval by nonrigid image matching, *Proc. 16th ACM international conference on Multimedia* (*MM 2008*), pp.41–50 (2008).

(平成 21 年 12 月 19 日受付)
(平成 22 年 4 月 6 日採録)

(担当編集委員　牛尼 剛聡)

山本　祐輔（正会員）

京都大学大学院情報学研究科博士後期課程在学中．日本学術振興会特別研究員（DC2）．2006 年京都大学工学部情報学科卒業．2008 年京都大学大学院情報学研究科修士課程修了．情報の信憑性，ウェブマイニング，研究室運営に関する研究に従事．人工知能学会，日本データベース学会各学生会員．

田中　克己（正会員）

京都大学大学院情報学研究科社会情報学専攻教授．1976 京都大学大学院修士課程修了．博士（工学）．主にデータベース，マルチメディアコンテンツ処理の研究に従事．IEEE Computer Society，ACM，人工知能学会，日本ソフトウェア科学会，日本データベース学会等各会員．