

Web上の画像・テキスト対の信憑性分析

山本 祐輔[†] 田中 克己[†]

[†] 京都大学大学院 情報学研究科 社会情報学専攻

E-mail: †{yamamoto,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 本稿では Web 上に存在する画像・テキスト対の信憑性を評価するモデルを提案する。本稿で着目する信憑性は画像とそれを説明しているテキスト間の整合性である。提案モデルでは画像・テキスト対の整合性を分析するため、関連する画像・テキスト対集合内における関係性を分析することで、(1) テキストに対する画像の典型性、(2) 類似画像を説明するテキスト間の無矛盾性の評価を行い、その上で (3) 画像典型性とテキスト間の無矛盾性を考慮した画像・テキスト対の整合性を評価する。本稿ではまず、関連する画像・テキスト対が全て得られる理想的な状況下での信憑性評価方法をモデル化する。次に、ドメインを限定し検索エンジンを用いた現実的な画像・テキスト対の信憑性分析方法について述べる。提案アプローチによって検索エンジンベンダー側は信憑性の低い情報を予め排除する、ユーザは情報閲覧中に画像・テキスト間の信憑性に疑問を抱いたときに検証を行う、といったことが可能となる。

キーワード Web 情報の信憑性, 画像・テキスト対の信憑性, 画像解析, 2 部グラフ問題



ハンバーガー H は今までのハンバーガーよりもすごい!



実際ハンバーガー H は全然違った。こんなべったこじゃないか!

(a) 広告上のハンバーガー H の写真 (b) ブログ上のハンバーガー H の写真

図 1 広告画像が実際の商品を食べたユーザが掲載した画像と異なっているケース。広告 (a) は見た目の良いハンバーガー写真を用いることでハンバーガー H の良さを強調している。一方でブログ (b) で掲載されているハンバーガー H の写真は (a) の姿とは異なりボリュームが少ない。

1. はじめに

今日ユーザは存在する膨大な Web 情報から必要な情報を容易に取得できる。Web には有益な情報も多く存在する反面、誰しもが Web に情報を公開できることから信憑性の低い情報も多く存在する。Web の価値判断は完全にユーザに依存しているので、信憑性の低い情報を知らず知らずのうちに鵜呑みにしてしまうケースが存在する。したがって情報の信憑性を何らかの形で評価、判断支援する機構が必要となる。

本研究では Web 情報の中でも特に画像とその説明テキストのペアに焦点を当てた信憑性について論じる。画像情報はテキストのみの情報に比べて圧倒的に内容の理解が容易であるため、画像およびそれに関連する説明テキストの信憑性が低い場合、大きな問題となる。例えば、図 1 は Web 広告に掲載されたあるハンバーガー H の画像と、それを食べたと主張するユーザが書いたブログに掲載された同じハンバーガー H に関する画像



Web ページ A

足利尊氏は征夷大將軍となり、京都に幕府を作った。



Web ページ B

この肖像画の人物は足利尊氏と言われていたが、近年は尊氏の従者であった高師直という説が強い。

図 2 同一画像に対して主張が異なっているケース。Web ページ A では肖像画の人物を足利尊氏と言及しているのに対して、Web ページ B では高師直と言及している。

である。ハンバーガー H を説明するために用いられている画像に食い違いがあるため、どちらのハンバーガー H の画像が実物に近い典型的な商品を表しているかは分からない。仮に多くの Web ページに図中の (b) と同様の写真が掲載されている場合、(a) の広告はハンバーガー H の典型的な姿を現していないという意味で信憑性が低い画像を掲載していることになる。この例はあるラベルに対応する画像の信憑性に関する問題を示唆している。また図 2 は馬に乗った侍に関する同一の画像に対して Web ページ A では画像の人物を「足利尊氏」と説明しているのに対して、Web ページ B では「高師直」と説明している例である。この例では同じ画像に対して人物名という本来背反するものが説明テキストに記述されており、画像に対する説明テキストの信憑性に関する問題を示唆している。

このように画像・テキスト対の信憑性といっても、テキスト内容に対する画像の分析と画像に対するテキスト内容の分析の 2 方向が考えられる。しかし、画像・テキスト対の信憑性を評価するためには、片方向からのみの評価では不十分であり、両方向からの評価を同時に行う必要がある。我々はこれまでラベルに対する画像の信憑性を画像の典型性から評価する手法を提

案してきたが[1], この手法ではしばしば典型的な画像が実は誤りであるというケースが生じてしまう. 例えば, Google 画像検索で「足利尊氏」というクエリで検索を行うと図2のような騎馬武者の画像が多数得られる. そのため, 人物名に対する画像の典型性評価だけでは実際には別人物の写真であることが指摘されているにも関わらず誤った判断をしてしまう可能性がある. すなわち, テキストに対する画像の典型製水かだけでは, 画像・テキスト間の整合性を判定できない. そこで本研究では, 画像・テキスト対の信憑性を**画像・テキスト間の整合性の問題**として捉え, テキスト内容が指している画像の典型性, および類似画像に付与されている説明テキスト間の無矛盾性, の両要素を考慮して分析する手法を考える.

本稿では, 任意の画像・テキスト対の信憑性の分析に関連する画像・テキスト対集合を分析することによって行う. ここである画像・テキスト対に関連する画像・テキスト対とは, 対象画像と画像が類似している画像・テキスト対, 対象テキストとテキスト内容が類似している画像・テキスト対, あるいは両方を満たすものである. 関連画像・テキスト対集合と比較することで画像・テキスト間の統計的な整合性から信憑性を評価することを旨とする. そこで, まず第2.章で画像・テキスト対の信憑性分析を2部グラフ上でモデル化する. 第3.章では, ドメインを限定して画像・テキスト対の信憑性分析を限られたリソース下で分析する上での方法を提案しアルゴリズムの利用例を示す. 第6.章では信憑性に関する関連研究との比較を行い, 最後に本研究の課題を述べる.

2. 画像・テキスト対の信憑性分析モデリング

本章では, Web 上に存在する画像・テキスト対の信憑性およびその分析手法について2部グラフを用いてモデリングを行う. 前章で述べたように, 本稿で扱う画像・テキスト対の信憑性は画像・テキスト間の整合性として捉え, 関連する画像・テキスト対集合との比較によって評価される.

画像 i とその説明テキスト t との整合性が高いケースでは, 関連する画像・テキスト対の集合との比較で考えると, 直感的にはテキスト t を説明するために他の画像に比べ画像 i が用いられることが多く, 同時に画像 i に対してはテキスト t の内容で説明されることが多いと考えられる. つまり画像 i とテキスト t の対の整合性 $Cred(t, i)$ は

$$Cred(\langle i, t \rangle) \propto Pr(i|t) \quad (1)$$

$$Cred(\langle i, t \rangle) \propto Pr(t|i) \quad (2)$$

のように表現することができる. よって最終的には上式を満たすような関数 f :

$$Cred(\langle i, t \rangle) = f(Pr(i|t), Pr(t|i)) \quad (3)$$

を定義することによって画像・テキスト対 $\langle i, t \rangle$ を評価することになる.

実際にはテキスト t の内容を表す画像と全く同一の画像 t が

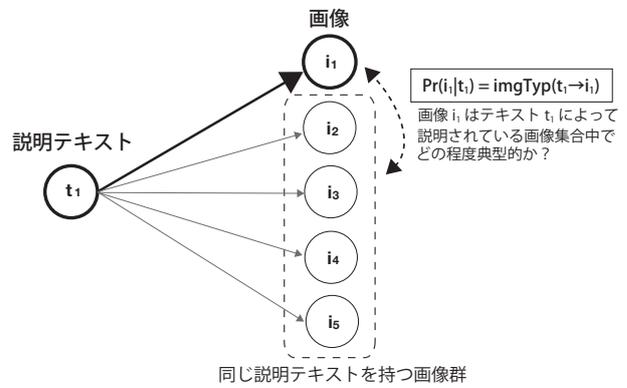


図3 ある説明テキストに対する画像の典型性評価. 太線で描かれた画像・テキスト対が評価対象.

複数存在するということはありません, $Pr(i|t)$ を単純に計算することはできないので, $Pr(i|t)$ はテキスト t を説明するために画像 i とよく似た画像が用いられている確率を考えるべきである. $Pr(t|i)$ についても同様に画像 i とその類似画像を説明するのにテキスト t が用いられている確率として考える.

ところで, ある文章内における画像の用いられ方は本来は多様であり, 画像の内容をテキストで説明するケースもあれば, テキストの内容の補完や内容の視覚化のために画像が用いられるケースもある [2], [3]. 今回画像を説明しているテキストとして問題としているのは, 同一 (類似) 画像上で本来同時に述べられるべきでない背反したテキスト内容が述べられているケースである. 例えば仮に図2の Web ページ A の類似画像に「貴重な肖像画を博物館で見た」というテキストが付けられていたとしても「足利尊氏」とテキストで説明された画像と食い違いの対象とはならない. つまり $Pr(t|i)$ は画像 i およびそれと類似している画像に対してテキスト t と比較して背反しているテキストによって説明されているケースの2クラスの画像・テキスト対集合の中で計算されるべきである.

ある画像・テキスト対の信憑性分析を行うためには, 関連する全画像・テキスト対を収集する必要がある. また, 実際には画像は Web 文書内で用いられるが文書内から画像の説明として対応している箇所を正確に抽出しなければならない. 本章では, 分析に必要な画像・テキスト対は何らかの方法で収集でき, 画像に対する説明テキストが同定できるという理想的な状況を考える.

2.1 あるテキストに対する画像の典型性

説明テキスト t に対して画像 i が使用される確率が高いというのは, 直感的には説明テキスト t に対して画像 i と類似する画像が頻繁に使用されている, テキスト t に対して画像 i が典型的な画像として使用されていると言い換えることができる. よって, テキスト t に対する画像 i の整合性を単純に評価するには, 図3のように, テキスト t を同じ説明テキストとして持つ画像集合を取得し, その集合内における画像 i が視覚的な典型性 $imgTyp(t \rightarrow i)$ を評価すればよい. このことからテキスト t と画像 i の整合性 $Cred(\langle i, t \rangle)$ は以下のように表現することができる.

$$Cred(\langle i, t \rangle) \propto Pr(i|t) \quad (4)$$

$$\approx imgTyp(t \rightarrow i) \quad (5)$$

$$= imgTyp(i, DB(txt = t)) \quad (6)$$

$DB(txt = t)$ は画像データベースに存在する画像でテキスト t を説明テキストとして持つ画像集合を表し, $imgTyp(i, Images)$ は画像集合 $Images$ 内における画像 i の典型性を表す。

実際にテキストに対する画像の典型性 $imgTyp$ を求めるには, テキスト t を説明テキストとしてもつ画像群 $DB(txt = t)$ と画像 i を画像特徴量でクラスタリングして, 画像 i が属するクラスターの $DB(txt = t)$ 集合を占める割合を計算する方法, 画像 t の $DB(txt = t)$ 集合内における視覚的な中心性を計算する方法などが考えられる。本稿では, 後者のアプローチを採用し, それを実現するために VisualRank アルゴリズムを用いた [4], [5]。

VisualRank アルゴリズムは画像集合が与えられたときに, 全ての画像間に画像間類似度を重みとする枝を張ることでグラフ構造を作成し, それに対して PageRank アルゴリズムを適用することで与えられた画像集合内の各画像の視覚的な中心性を求めるものである。画像集合が与えられたとき全画像間の類似度を表す類似度行列を S^* , PageRank における *damping factor* を d , p を *damping vector* としたとき, 各画像の画像集合中での視覚的な中心性を表すベクトル VR は以下の数式により再帰的に計算される：

$$VR = dS^* \times VR + (1 - d)p, \text{ where } p = \left[\frac{1}{n}\right]_{n \times 1} \quad (7)$$

本稿では, この式で計算された VR の各画像に対応する値を $imgTyp(t \rightarrow i)$ として利用する。

2.2 ある画像に関するテキスト内容の無矛盾性

2.1 章で述べた方法は「テキストから画像」という方向の妥当性を評価する方法であったが, この方法では図 2 のように類似した画像に背反するような説明テキストが張られているケースには対応できない。一般に文書内で画像が用いられるコンテキストは様々であるため, 画像に対応する説明テキストが他種類存在することがありえる。類似画像に張られている説明テキストが評価したい画像・テキスト対のテキスト内容と関連性が無く背反関係でもない場合, その影響を無視すればよい。一方で背反関係になっているような場合は, ある類似画像群上の異なったテキスト内容間で矛盾が生じていることを考慮する必要となる。画像 i に張られている説明テキスト t の無矛盾性 $txtCons(i \rightarrow t)$ を評価するには, 直感的には図 4 のように, 画像 i と類似している画像類似画像を収集したときに, テキスト t と同様の内容のテキストと背反する内容のテキストの 2 クラスの出現頻度を比較すればよい。画像 i に張られている説明テキスト t の無矛盾性を $txtCons(i \rightarrow t)$ とすると, 画像 i とテキスト t の整合性 $Cred(\langle i, t \rangle)$ は次のように表現することができる：

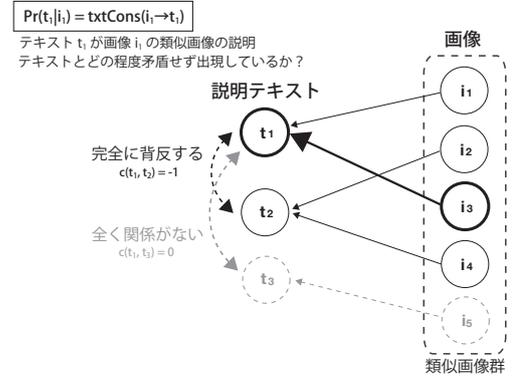


図 4 ある類似画像に対する説明テキストの無矛盾評価

$$Cred(\langle i, t \rangle) \propto Pr(t|i) \quad (8)$$

$$\approx txtCons(i \rightarrow t) \quad (9)$$

$$= txtCons(t, \{t_k | i_k \in DB(img = i), \langle i_k, t_k \rangle\}) \quad (10)$$

ここで $DB(img = i)$ は画像データベースから収集された画像 i の類似画像集合を表し, $txtCons(t, Texts)$ はテキスト集合 $Texts$ 内でテキスト t とそれと類似するテキストがテキスト t と背反するテキスト集合と比べてどの程度存在するかを表す数値である。

ある説明テキストがテキスト t と類似した内容か矛盾した内容かを判定する際, 扱うテキストが商品名やランドマーク名といったオブジェクト名に注目した場合はテキスト t 以外のテキストは全て背反していると考えられる。一方で, 説明テキストを単語群または自然言語のようなものであった場合, 内容がやや類似しているもの, 背反しているものも存在する。このような問題に対して, テキスト t と各説明テキストの間に背反度 c を導入することで対応することができる。今あるテキスト間の背反度を $-1 \leq c \leq 1$ とし 1 に近いほどテキスト間の内容的に類似し, -1 に近いほど矛盾していると定義する。今,

$$|SimilarTxt(i)| = \sum_{\langle t_x, i_x \rangle | c(t_x, t) \geq \theta, i_x \in DB(img=i)} c(t_x, t) \quad (11)$$

$$|ConflictTxt(i)| = \sum_{\langle t_x, i_x \rangle | c(t_x, t) < \theta, i_x \in DB(img=i)} -c(t_x, t) \quad (12)$$

とすると, 画像・テキスト対 $\langle i, t \rangle$ の整合性 $Cred(\langle i, t \rangle)$ は以下のようなになる。

$$Cred(\langle i, t \rangle) \propto txtCons(i \rightarrow t) \quad (13)$$

$$= \frac{|SimilarTxt(i)|}{|SimilarTxt(i)| + |ConflictTxt(i)|} \quad (14)$$

ここで $c(t_x, t)$ はテキスト間の背反度, θ は閾値 (≥ 0) とする。

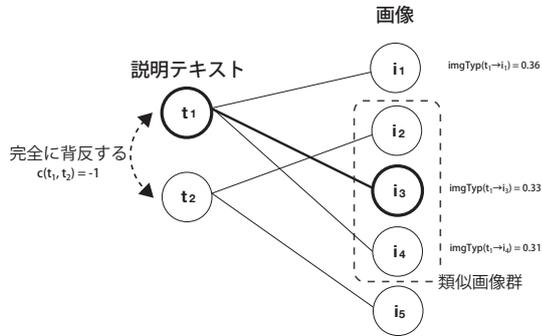


図5 画像とテキストの関係性の一例

2.3 画像・テキスト対の信憑性

これまで述べてきたテキストに対する画像の典型性および画像に対するテキストの無矛盾性の両方を用いて、画像・テキスト対の信憑性の評価モデル (3) を以下のように表す：

$$Cred(\langle i, t \rangle) = f(\text{imgTyp}(i \rightarrow t), \text{txtCons}(t \rightarrow i)) \quad (15)$$

$$= \text{imgTyp}(i \rightarrow t)^\alpha \cdot \text{txtCons}(t \rightarrow i)^\beta \quad (16)$$

α, β は画像典型性、テキスト無矛盾性の影響力をコントロールするパラメータである。これらのパラメータをどのように設定するかはどのような画像・テキスト対の信憑性を評価するかによって異なる。例えば、図1のようにあるオブジェクト名に対して典型的な画像が選択されているかを評価したい場合はパラメータ β を0にセットすればよい。また、図2のようにある画像を記述しているテキストに矛盾があるかないかを評価したい場合はパラメータ α を0にセットする。あるテキスト内容に対して矛盾が無く典型的に用いられる画像を収集したい場合は画像典型性とテキスト無矛盾性の両方を考慮するようにパラメータ設定すればよい。画像典型性に対してテキスト無矛盾性を乗算することは、**背反するテキスト内容を持つ画像の投票効果を相殺する、あるいは背反しているテキストが存在しているときに罰則を与える演算**であると考えられる。なお $\text{txtCons}(t \rightarrow i)$ を計算するには図5のような状況を図6のように類似画像を集約する。そのため $\text{imgTyp}(i \rightarrow t)$ の計算の対称性を保つためにテキスト t を説明テキストとしてもつ画像のうち画像 i と類似している画像・テキスト対の imgTyp の平均値を画像典型度 $\text{imgTyp}(t \rightarrow i)$ として用いる。

3. 実装

これまで述べてきた画像・テキスト対の信憑性分析モデルはWeb上に存在する全ての画像が得られ、さらに画像に対する説明テキストも完全に同定されていることが前提であった。しかし、Web上の全ての画像にアクセスすることは事実上困難であり、また画像と説明テキストの対応付けも明示的にメタデータが付与されていないようなデータに対しては困難である。そこで、本章では提案手法が適応できるドメインとして歴史上人物

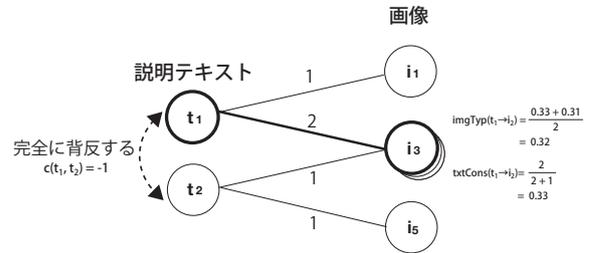


図6 画像典型性とテキストの無矛盾性を考慮した画像・テキスト対の信憑性分析。枝に描かれた数はその画像・テキスト対が何個存在するかを示している。

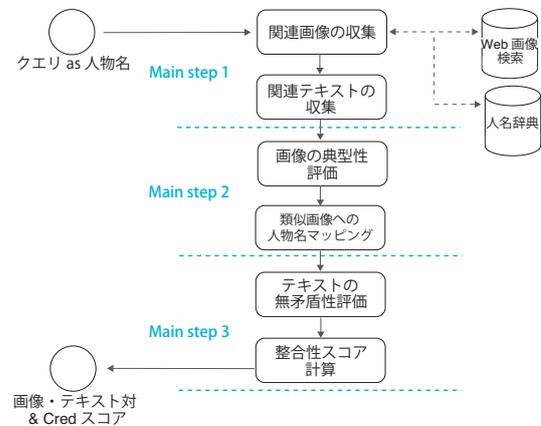


図7 画像・テキスト対の信憑性分析の流れ

画像と人物名の画像・テキスト対に限定することで現実的な信憑性分析システムの実装について述べる。

提案システムは入力として歴史人物名を受け取ると、典型的でかつ人物ラベルに矛盾が起こっている可能性が低い画像をランキングして返す。分析の手順は大きくは3つある。ステップ1では、クエリと既存の画像検索エンジンを用いてクエリと関連性のある画像およびそれらの画像上で背反している可能性がある別の人物名を抽出する。ステップ2では、画像解析を行い入力人物名に対する画像の典型性を評価し、さらに類似画像に対して異なる人物名が割り振られていないかの確認を行う。ステップ3では、画像に対する人物名の無矛盾性を評価し、その上で各画像・テキスト対の信憑性を整合性という観点から評価する。図7はシステムの流れを示している。

3.1 画像・テキスト対の収集

入力された人物名に関する画像・テキスト対の整合性分析を行うためには、まず入力された人物名に関する画像・テキスト対および類似画像上で別人物の名前が割り当てられている画像・テキスト対をまず収集する必要がある。

ある画像に対して割り当てられた人物名が誤っている場合、その誤りを正すために別人物の名前が画像の周辺テキストに出現していることが考えられる。そこで、まず入力クエリを用いて画像検索を行い、収集された画像集合の周辺テキストから人物名を抽出する。画像の周辺テキストを取得する方法はいくつか提案されている [6], [7]。本稿では画像の前後文章から50語を周辺テキストとして抽出する。そしてその周辺テキストに出現する人物名を抽出する。この際、事前に用意した7250人の

人物名が掲載された歴史人物辞書を使い、歴史人物名だけを抽出する。

初期クエリから人物名を収集後、各人物名を再度 Web 画像検索エンジンに投げ、各検索結果の上位 N 件の画像を取得する。人物クエリ q で得られた画像集合は人物名 q が割り当てられていると考え、全ての画像に対して人物名を割り振り、解析用の画像・テキスト対を作成する。

3.2 画像典型性の分析と類似画像に対する人物名マッピング

ステップ 2 では、まず入力された人物名に対応する全画像の典型性を評価する。ある人物名に対する画像の典型性を評価するには、2.1 章で述べたように画像間の視覚的な類似度を計算し、VisualRank アルゴリズムを適用する。画像間の類似度を計算する方法は多数存在するが、大きくは global-feature に基づく分析と local-descriptor に基づく分析に分けられる。global-feature に基づく分析としては色ヒストグラム類似度 [8]、テクスチャ類似度 [9]、離散コサイン変換 (DCT) による色と形状を同時に考慮した類似度計算 [10] などが良く用いられる。しかし、global-feature は対象としている画像の種類に影響を受けやすいので類似度計算が難しい。一方で local-descriptor は画像のスケールや回転などに影響を受けず画像の特徴を表現することができる [11], [12]。本稿では local-descriptor である Scale Invariant Feature Transform (SIFT) を用いて画像間の類似度計算を行う [13], [14]。各画像に SIFT を適用すると 128 次元のベクトルで表現された画像特徴点が幾つか抽出される。本稿では、画像 i_a と i_b が与えられたとき、共通する特徴点の数を含む割合で画像間の類似度を定義する。具体的には、画像 i_a , i_b の特徴点の数をそれぞれ $Key(i_a)$, $Key(i_b)$ とし、共通する特徴点の数を $SharedKey(i_a, i_b)$ としたとき、2 枚の画像間の類似度 $Sim(i_a, i_b)$ は、

$$Sim(i_a, i_b) = \frac{1}{2} \left(\frac{SharedKey(i_a, i_b)}{Key(i_a)} + \frac{SharedKey(i_a, i_b)}{Key(i_b)} \right). \quad (17)$$

と定義する。この類似度の定義を用いて入力人物名に対する画像の典型性を評価する。

また画像間の類似度を計算することで、入力人物名に対応する画像に別人物の名前が対応づけられているケースを発見する。入力人物名に対応する各画像と他人物の全画像との類似度を計算し、閾値 (実験では 0.1 と設定) を超えた場合、他人物の名前をその画像に対応づける。このような作業により図 5 のような 2 部グラフを構築する。

3.3 画像・テキスト対の整合性分析

全ステップで類似画像分析によって入力人物名 q_{input} に対応する画像に他の人物名が割り振られているかどうかの判定が済んでいる。今人物名 q_{input} に対応する画像集合を $Images(q_{input})$ としたとき、画像 $i_k \in Images(q_{input})$ に割り振られている入力人物名 q_{input} の無矛盾性を評価する。今回は説明テキストを人物名と限定しているので、入力人物以外の人物名が出現している場合は「完全に背反している」と見なす。つまり、式 (11), (12) における背反度 c は 1 か -1 しか取りえない。このような

設定で各画像の入力人物名に対する無矛盾性を評価する。

入力人物名に対する画像の典型性と画像に対する人物名の無矛盾性を評価した後に、最終的に式 (16) を用いて画像・テキスト対の整合性を評価する。実際のシステムでは無矛盾性をより重視した整合性を評価するために $\alpha = 1$, $\beta = 3$ とする。

図 8 は提案アルゴリズムを画像検索結果の上位 50 件のリランキングに適用した例である。ここでは入力「足利尊氏」という人物名が与えられたときの検索結果の違いが分かる。図中の (a) が Google 画像検索を行った上位 15 件の結果であり、(b) は提案アルゴリズムの検索結果上位 15 件である。(b) の検索結果には $Cred(\text{足利尊氏}, i)$ のスコアと、類似画像上で背反する人物名が出現している場合にはその人物名が提示されている。既存の検索結果では騎馬武者の画像が多く見られるのに対して、提案アルゴリズムでは別人物であるといわれている騎馬武者 ((a) の画像 1,2,3,6,8,13,15), 平重盛 ((a) の画像 5), 源頼朝 ((a) の 14) はランキング下位に下がり、現時点での歴史研究で「足利尊氏である」と分かっている像に関する画像などがランキング上位に上がってきている。提案アルゴリズムでも騎馬武者に関する画像がランキング上位に残っているが、類似画像に割り振られている人物名が提示されているので、ユーザは違う人物であることを意識しながら画像を検索することができる。

4. 実験

提案手法の有効性を確かめるために 3 種類の実験を行った。実験では画像・テキスト対の分析対象として、歴史上の人物とそれを指している画像の対を選んだ。

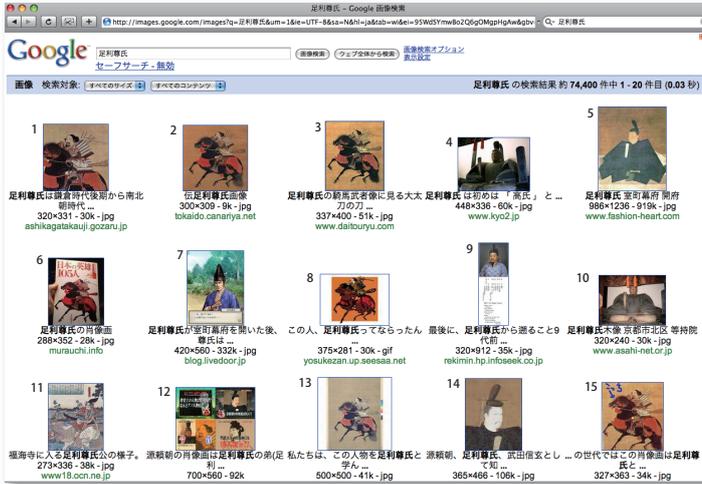
4.1 実験設定

今回の実験では 3. 章で述べたアプローチに従って画像・テキスト対の整合性を評価した。画像収集のための検索エンジンには Google 画像検索エンジン^(注1)を用い、検索結果上位 50 件を解析対象とした。提案モデルではより多くの画像を解析対象にした方が整合性評価の妥当性が向上すると考えられるが、今回のアプローチでは検索エンジンを使っており、検索結果が下位になればなるほどノイズが含まれることが多くなる。よって今回 50 件という検索結果数に設定した。画像間の類似度の計算をする際には、実際の画像ではなく Google が提供しているサムネイル画像を対象にした。画像典型性 $imgTyp$ の計算に用いる damping factor p は 0.85 に、テキストの無矛盾性 $txtCons$ を計算する際の画像クラスタリングの閾値 θ は 0.1 とし、また画像・テキスト対の整合性 $Cred$ の計算に用いるパラメータは $\alpha = 1$ および $\beta = 3$ と設定した。

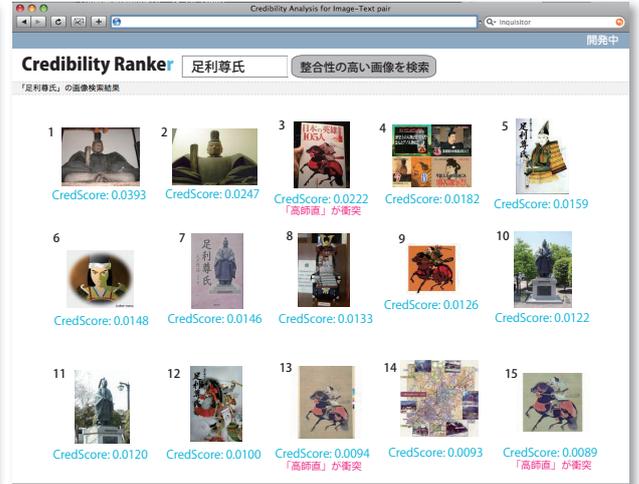
4.2 整合性の低い画像・テキスト対に対する評価能力

提案手法を評価するために、まず提案アルゴリズムが不整合な画像・テキスト対を「不整合である」と判定できるかを評価した。この評価を行うために、ある画像に関して、既に歴史研究によって別人の可能性があると分かっている歴史人物 6 人を選択した。選択された 6 人は、「足利尊氏」「高師直」「足利直義」「武田信玄」「平重盛」「源頼朝」である。

(注1) : Google 画像検索, <http://http://images.google.co.jp/>



(a) Google 画像検索における「足利尊氏」の結果上位 15 件



(b) 提案システムにおける「足利尊氏」検索の結果上位 15 件

図 8 既存の画像検索エンジンの結果と提案アルゴリズムの分析結果の差異

実験では、これらの人物名をクエリとして得られた画像検索の上位 50 件を解析対象とした。そして得られた画像集合に対して提案アルゴリズムを適用し整合性順にランキングを行った。評価はランキングの上位 3,5,10 件以内に実際に不整合な画像が何件あるかを、Google 画像検索 (ベースライン)、VisualRank アルゴリズム、提案アルゴリズム毎に比較した。

図 9 は上位 N 件毎の比較結果である。図によると提案アルゴリズムは他のアルゴリズムよりも多くの不整合画像を排除できている。評価範囲を上位 10 件以内に拡大してもほとんどの不整合画像を排除できていることから、提案手法が不整合画像の排除に有効であることが分かる。上位 10 件以内に限定すれば、提案アルゴリズムでは 8.3%の不整合画像を含むのに対し、Google 画像検索では 45%、VisualRank アルゴリズムでは 66.7%の不整合画像を含んでいた。

図 9 の結果から、提案アルゴリズムは有効に機能しているように見えるが、本来整合性の高い画像まで低く評価している可能性も否定できない。そこで、整合性の低い画像に加え、人物名に対して適合性の低い画像をどの程度含んでいるかを評価した。その結果を図 10 に記す。図によると、提案アルゴリズムは不整合な画像を含んでしまっているが、他の 2 つのアルゴリズムよりも整合性の低い画像および適合性の低い画像を排除できている。上位 20 件以内に限定すると、提案アルゴリズムは 51.7%の不整合および不適合画像を含んでいた。画像・テキスト対の信憑性に焦点を当てると、不整合な画像はそれを含む文書をユーザが閲覧すれば不整合であると確認できることから、不整合な画像よりも不整合な画像を排除することが重要である。それゆえ提案アルゴリズムは今回の別人の可能性が疑われる画像を持つ歴史人物に対しては有効に働いたと言える。

4.3 整合性の高い画像・テキスト対に対する評価能力

提案アルゴリズムが本来整合性の高い画像・テキスト対を誤って不整合と判定しないかを評価するために、別の実験を行った。この実験では明らかに整合性が取れていると事前に分かっている肖像画を持つ歴史人物 6 人を選び、各歴史人物名で画像を取

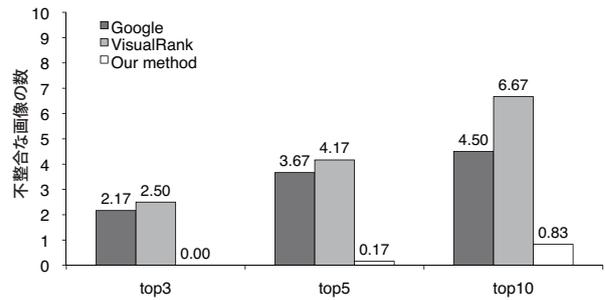


図 9 上位 N 件に含まれる不整合画像の平均数の比較

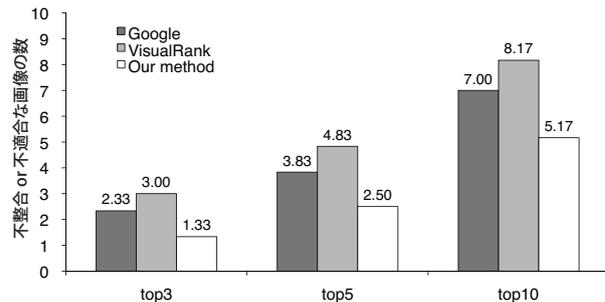


図 10 上位 N 件に含まれる不整合画像と不適合画像の平均数の比較

集し、提案アルゴリズムでランキングしたときに上位 3,5,10 件に整合性のある画像をどの程度含むかを評価した。評価は全実験と同様 Google 画像検索、VisualRank アルゴリズムと比較した。実験に用いた歴史人物名は「足利義満」「織田信長」「豊臣秀吉」「徳川家康」「坂本龍馬」「大久保利通」である。図 11 に結果を記す。

図によると、提案アルゴリズムは他のアルゴリズムに比べて整合性の高い画像をランキング結果上位に含むことができなかったが、結果にそれほど大きな差は無かった。上位 10 件に限れば、提案アルゴリズムの返す結果には平均 8.67 枚の整合性の取れた画像を含むことができた。整合性の高い画像に対し

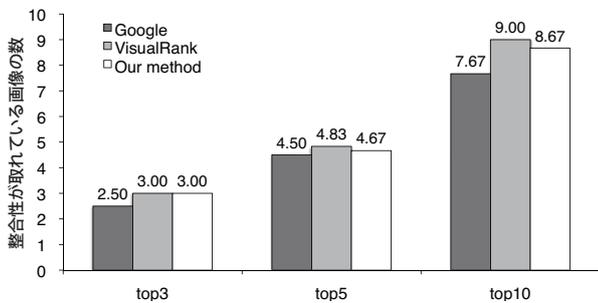


図 11 上位 N 件に含まれる整合性が取れている画像の平均数の比較

ては、他のアルゴリズムと同程度画像を評価することができたとと言える。

5. 考 察

提案モデルでは解析対象となっている画像・テキスト対に関連する全ての画像・テキスト対が漏れなく収集できるという理想的な環境を想定した。さらに、画像・テキスト対上では画像とテキスト間の対応が保証されていることを想定していた。しかし、実際にはこの理想状況を実現することは難しく、画像検索エンジンを用いて Web 上の関連画像を収集することしかできないのが実情である。検索エンジンを用いた場合の弊害としては、あるクエリで画像を収集した際、クエリであるキーワード(群)が本文中で画像を指しているとは限らない。これは既存の画像検索エンジンはあくまで画像の alt タグや周辺テキストにクエリが出現するか否かで画像を検索しているためである。また、類似画像上に対して背反するテキストが張られていないかを検証するため、テキストベースの画像検索エンジンを用いているが、本来はコンテンツベースの画像検索によって類似画像を集めるべきである。これらの問題は画像検索エンジンの検索精度に関わる問題であり、これらが改善することで提案モデルの適用が現実的になるであろう。

提案アルゴリズムが画像・テキスト対の整合性を評価する上での最大の特徴は同一(類似)画像上で背反する内容のテキストが存在する場合を考慮することである。今回は人物名の背反関係に焦点を当てたが、ここで意味する背反とは同時に起こりえるはずのない事象・命題に着目した論理的な背反である。一方で同一の画像に対して「札幌時計台は素晴らしい」と「札幌時計台にはがっかり」と言ったような意見・印象の背反も起こりえる。今回提案したモデルでは背反度 c が導入されているため、この値の定義によっては後者の背反性についても扱えるが、意見・印象が同一の画像上で対立していることが画像・テキスト対の信憑性、特に整合性の問題として重要かどうかは十分な考察が必要である。

6. 関連研究

情報の信憑性の評価および判断支援方法に関する研究は近年幾つか行われている。Fogg らはユーザ視点から見た Web サイトの信憑性を測るための尺度を体系化し、各尺度をユーザがどの程度意識しているかを大規模な調査により明らかにしてい

る [15]~[17]。しかし、尺度の整理はされているものの、各尺度に関して具体的な評価方法およびアプリケーションが提案された先行研究は未だ少ない。筆者らはユーザが「納豆はダイエットに効果的か?」といった不確かなファクト型知識の信憑性の判断するための材料を Web 情報の集約によって提示するシステムを提案している [18], [19]。

画像を信憑性検証の観点から評価するという点では画像検索のランキング技術に関連があり、これらの関連研究は非常に多い [20]。これらの研究の目的は、ユーザが入力したクエリに適合する画像を収集することであるが、適合率はあくまでキーワードと画像がどの程度共起するかの指標であるため、本稿で述べたような「足利尊氏」の画像のように適合性が高くとも、その画像がクエリを表しているとは限らない。本稿で提案したアルゴリズムは信憑性ベースの画像検索アルゴリズムに適用できる可能性がある。

7. おわりに

本稿では、画像・テキスト対の信憑性を分析するために、関連する画像・テキスト対集合を統計的に分析することで評価するモデルを提案した。提案モデルでは画像・テキスト対の信憑性を、(1) テキストに対する画像の典型性、(2) 画像を説明しているテキストの無矛盾性、(3) 画像の典型性およびテキストの無矛盾性の両要素を考慮した画像・テキスト対の整合性の 3 つの尺度からモデル化した。検索対象の種類、ユーザの意図など、アプリケーションに提案した 3 つの尺度を使い分けることがより有用であると考えられる。

画像・テキスト対の信憑性を評価するためには提案モデル以外にも様々なモデルが考えられる。提案手法では画像・テキスト対集合を統計的に処理することで注目している画像・テキスト対の整合性を評価しようと試みたが、画像・テキスト対が掲載している Web ページの発信者の特性を考慮した画像・テキスト対の評価などが考えられる。画像・テキスト対の整合性以外にも、画像特有の信憑性として画像内容の鮮度の問題など様々な問題が考えられる。信憑性は分析対象である情報の種類、ユーザの情報を閲覧するコンテキストなどによっても異なるため、1 つの評価尺度にだけでなく複数の有用な評価尺度から信憑性分析を行うことが重要であると考えられる。

最後に、今回提案した手法はユーザがある画像の信憑性を疑っているときに判断材料を探すシステムとしては有効であるが、本来信憑性の問題は、ユーザが対象の信憑性を全く意識していないときに重要になる。今回提案した判断材料の提示システムを改善することと併せて、信憑性が問題がある画像を閲覧している際にそれを通知する機構の開発も必要である。今後は、本稿で提案した手法の大規模な定量的評価を行う予定である。

謝辞 本研究は一部、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」、科研費：計画研究「情報爆発時代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」(課題番号 18049041)、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)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] 山本祐輔, 山本岳洋, 中村聡史, 田中克己, “関連画像集合内における典型度と特殊度を用いた画像の信憑性分析,” Webとデータベースに関するフォーラム (WebDB Forum 2008) 予稿集, pp.inCD-ROM, 2008.
- [2] J.H. Larkin and H.A. Simon, “Why a Diagram is (Sometimes) Worth Ten Thousand Words,” *Cognitive Science*, vol.11, no.1, pp.65–99, 1987.
- [3] G.L. Lohse, K. Biolsi, N. Walker, and H.H. Rueter, “A classification of visual representations,” *Commun. ACM*, vol.37, no.12, pp.36–49, 1994.
- [4] Y. Jing and S. Baluja, “Pagerank for Product Image Search,” *Proc. of the 17th international conference on World Wide Web (WWW’08)*, pp.307–316, 2008.
- [5] Y. Jing and S. Baluja, “VisualRank: Applying PageRank to Large-Scale Image Search,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.30, no.11, pp.1877–1890, 2008.
- [6] D. Cai, X. He, Z. Li, W.-Y. Ma, and J.-R. Wen, “Hierarchical clustering of www image search results using visual, textual and link information,” *Proc. of the 12th annual ACM international conference on Multimedia (MULTIMEDIA 2004)*, pp.952–959, 2004.
- [7] Z. Li, S. Shi, and L. Zhang, “Improving relevance judgment of web search results with image excerpts,” *Proc. of the 17th international conference on World Wide Web (WWW 2008)*, pp.21–30, 2008.
- [8] M. Stricker and M. Orengo, “Similarity of color images,” *Proc. of IS&T/SPIE. Conference on Storage and Retrieval of Image and Video Databases III*, pp.381–392, 1995.
- [9] P. Howarth and S. Ruger, “Evaluation of texture features for content-based image retrieval,” *Proc. of the 3rd International Conference on Image and Video Retrieval (CVIR 2004)*, pp.326–334, 2004.
- [10] N. Ahmed, T. Natarajan, and K.R. Rao, “Discrete cosine transform,” *IEEE Transactions on Computers*, vol.23, no.1, pp.90–93, 1974.
- [11] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27, no.10, pp.1615–1630, 2005.
- [12] H. Chris and S. Mike, “A combined corner and edge detector,” *Proc. of the 4th Alvey Vision Conference*, pp.147–151, 1988.
- [13] D.G. Lowe, “Object recognition from local scale-invariant features,” *Proc. of the 7th International Conference on Computer Vision*, p.1150, 1999.
- [14] D.G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol.60, no.2, pp.91–110, 2004.
- [15] B.J. Fogg and H. Tseng, “The elements of computer credibility,” *Proc. of the 17th SIGCHI conference on Human factors in computing systems (CHI ’99)*, pp.80–87, 1999.
- [16] B.J. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, and M. Treinen, “What makes web sites credible?: A report on a large quantitative study,” *Proc. of the 19th SIGCHI conference on Human factors in computing systems (CHI 2001)*, pp.61–68, 2001.
- [17] B.J. Fogg, C. Soohoo, D.R. Danielson, L. Marable, J. Stanford, and E.R. Tauber, “How do users evaluate the credibility of web sites?: A study with over 2,500 participants,” *Proc. of the 2003 conference on Designing for user experiences (DUX 2003)*, pp.1–15, 2003.
- [18] Y. Yamamoto, T. Tezuka, A. Jatowt, and K. Tanaka, “Honto? search: Estimating trustworthiness of web information by search results aggregation and temporal analysis,” *Proc. of the 9th Asia-Pacific Web Conference (APWeb 2007)*, pp.253–264, 2007.
- [19] Y. Yamamoto, T. Tezuka, A. Jatowt, and K. Tanaka, “Supporting judgment of fact trustworthiness considering temporal and sentimental aspects,” *Proc. of the 9th International Conference on Web Information Systems Engineering (WISE 2008)*, pp.206–220, 2008.
- [20] T.A.S. Coelho, P.P. Calado, L.V. Souza, B. Ribeiro-Neto, and R. Muntz, “Image retrieval using multiple evidence ranking,” *IEEE Transactions on Knowledge and Data Engineering*, vol.16, no.4, pp.408–417, 2004.