ABSTRACT

Although the Web lets users freely browse and publish information, most Web information is unauthorized in contrast to conventional mass media. Therefore, it is not always credible or correct. We propose a model to solve these problems by enabling the credibility of text-image pairs on the Web to be analyzed. We propose a bipartite graph model, in which one set of nodes corresponds to a set of text data, and the other corresponds to a set of images. That is, each text-image pair is represented by an edge. We introduce the notion of “supportive relationships” among edges in our bipartite graph model. Intuitively, our hypothesis is that the more “supportive” text-image pairs a target text-image pair has, the more credible it is. Although such a bipartite graph model is by itself not necessarily new, one of its most notable features is that we take into consideration the similarity (or dissimilarity) among nodes in each node set to compute supportive relationships. As our model is generic, it can be applied to a variety of types of Web information represented by text-image pairs. We especially focus on the analysis of text-image pairs on the Web in this paper and describe a practical implementation of our model for analyzing credibility, ImageAlert.

Categories and Subject Descriptors

H.3.3 [Information Systems]: Information Search and Retrieval; I.4.9 [Computing Methodologies]: Image Processing and Computer Vision

General Terms

Algorithm

Keywords

Credibility of Web information, image credibility, image-text pairs analysis, image processing

*He also works as a Research Fellow (DC2) of the Japan Society for the Promotion of Science.

1. INTRODUCTION

Research aimed at evaluating the credibility of content on the Web has recently become even more crucial because it has started to influence our daily lives. The abundance of content on the Web, the lack of publishing barriers, and poor control of the quality of Web content has raised issues with credibility. Several researchers have reported that users perceive Web information to be somewhat credible [13, 14]. If users are not aware of the credibility of Web information, they can easily be misled, and this can sometimes endanger them. For example, there are more than twenty thousand health-related sites on the Web, but more than half such sites have not been reviewed by medical specialists [16].

Figure 1 shows horseback scenes of a certain Japanese historical personage on two Web pages. The portraits are of the same person in both these images, but their captions actually name two different people. A recent widely accepted interpretation actually explains that the person in these portraits is Kohno Moronao, not Ashikaga Takauji. Then, it follows that the caption on page A is incorrect.

Figure 2 shows another example of an exaggerated advertising illustration on the Web. For the same hamburger H, ad (a) exaggerates the value of hamburger H by using an attractive depiction and a catch phrase, while a consumer’s blog (b) indicates that the real hamburger H is awful unlike hamburger H in the ad. Thus, if captions of images provide incorrect facts like these, users are faced with the possibility of mistakenly identifying images. It is therefore important to ensure that images on Web pages are credible.

Three possible solutions are for users to check the credibility of images themselves, ask trustworthy people, or check for their credibility as exhaustively as possible from printed
books and Web-search engines. However, these are time-consuming and cannot always guarantee the credibility of pictorial content. Therefore, a system that allows the credibility of images to be estimated and that supports assessments of credibility is necessary so that these can be easily checked.

This paper focuses on the credibility of text-image pairs. This means that we checked for the correspondence or relationship between images and their text descriptions as previously mentioned and outlined in Figure 1. One possible approach to check for consistency between images and text is to identify objects in a single image, to interpret the content of the image, and to match it with the text content. However, this is not practical because a very large and exhaustive database needs to be prepared for images that are correctly labeled [4, 11, 19, 1]. Another possible approach is to check how frequently similar images appear when a text description is provided, and this approach is one conventional technique of image retrieval. However, even if some text descriptions semantically conflict and if text descriptions are frequently used with the image, this approach does not always guarantee the credibility of the target pair.

Therefore, we propose a bipartite graph model for analyzing the credibility of text-image pairs on the Web, in which one set of nodes corresponds to a set of text data, and the other corresponds to a set of images. That is, each text-image pair is represented by an edge. We introduce the notion of “supportive relationships” among edges in our bipartite graph model. Intuitively, our hypothesis is that the more “supportive” text-image pairs a target text-image pair has, the more credible it is. We introduce two basic concepts in this paper as the criteria to support analysis: edge typicality and edge peculiarity.

The credibility scores in our model are computed by analyzing a set of related text-image pairs that are gathered from the Web, not a single text-image pair. First, we introduce a basic model that enables the credibility of text-image pairs to be analyzed. After that, we introduce a practical algorithm for computing the credibility of text-image pairs focusing on a specific text description based on these criteria.

This paper makes three main contributions:

1. We model the relationship between texts and images from the viewpoint of credibility of text-image pairs by focusing on two basic aspects, i.e., edge typicality and edge peculiarity.

2. We propose a generic bipartite graph model for estimating the credibility of target text-image pairs by conducting a set-oriented analysis of the supportive relationships among it and its related text-image pairs. A notable feature of the proposed model is that we take into account the similarities (dissimilarities) among nodes in all node sets of a bipartite graph.

3. We focus on image-object name pairs as a simple case of text-image pairs, and we introduce a concrete method of evaluating the credibility between text-image pairs. In addition, we implemented a useful application, called IMAGEALERT, as one possible application that supports users to assess the credibility of a text-image pair on a Web page.

2. MODEL

2.1 Text-image pair (TIP) graph

Our target for analyzing credibility was an arbitrary text-image pair data appearing on the Web, such as a pair made up of an image and its label, and a textual statement and its associated image. It should be noted that we do not argue for the credibility of the target image itself as to whether or not it is real or faked in this paper. We only focus on the relationship between an image and its corresponding text (including image label).

To estimate the credibility of a target text-image pair, we search for and collect related pairs of the same type as the target from the Web and then analyze how “supportive” a text-image pair in a set of collected ones is of the target one. Finally, we estimate the credibility of the target text-image pair by analyzing how many supportive pairs it has.

A text-image pair graph (TIP graph, for short) is a bipartite graph \( G = (X, Y, E) \), where \( X \) is a set of text nodes, \( Y \) is a set of image nodes, and \( E \) is a set of edges, \( e = (x, y) \) \( (x \in X, y \in Y) \).

2.2 Supportive relations among edges

We created the following hypothesis to analyze credibility: The more numerous and more strongly supportive text-image pairs the target pair has in a set of related pairs, the more credible the target is.

There is an example of a target text-image pair and related pairs in Figure 3. Suppose that users want to know whether or not a target image is really a typical image that represents the given product name in this example. In other words, this focuses on the image’s typicality for the image label as the credibility criterion. The meaning of supportive text-image pairs for the target pair in this case is that images of the same (or similar) product are similar to the target pair’s image. To estimate the credibility of the target text-image pair in our approach, we check how many similar images with the same (or similar) product name there are on the Web. The target text-image pair in Figure 3 is regarded as having many similar text-image pairs. Therefore, we regard...
the target pair as receiving a great deal of support from the related pairs and conclude that it is credible.

For a given TIP graph, $G = (X, Y, E)$, and an edge, $e \in E$, the degree of credibility of edge $e$, denoted by $\text{cred}(e)$, is defined as:

$$\text{cred}(e) = \sum_{e' \in E} \text{sup}(e, e') \cdot \text{cred}(e')$$  \hspace{1cm} (1)

Here, $\text{sup}(e, e')$ denotes the degree of support of edge $e'$ for edge $e$, which will be defined later. Here, it should be noted that the degree of credibility, $\text{cred}(e)$, of edge $e$ is determined by the degree of support and the degree of credibility of edge $e'$.

Let $S$ and $C$ denote a matrix consisting of the degrees of support for each pair of edges and the degrees of credibility of all edges for $E = \{e_1, e_2, ..., e_n\}$ as:

$$S = \begin{bmatrix}
\text{sup}(e_1, e_1) & \text{sup}(e_1, e_2) & \cdots & \text{sup}(e_1, e_n) \\
\text{sup}(e_2, e_1) & \text{sup}(e_2, e_2) & \cdots & \text{sup}(e_2, e_n) \\
\vdots & \vdots & \ddots & \vdots \\
\text{sup}(e_n, e_1) & \text{sup}(e_n, e_2) & \cdots & \text{sup}(e_n, e_n)
\end{bmatrix} \hspace{1cm} (2)$$

$$C = \begin{bmatrix}
\text{cred}(e_1) \\
\text{cred}(e_2) \\
\vdots \\
\text{cred}(e_n)
\end{bmatrix} \hspace{1cm} (3)$$

Here, the matrix, $S^*$, is the column normalized adjacency matrix, where each $\text{sup}(e_i, e_j)$ denotes the degree of support $e_i$ to $e_j$. $C$ is the credibility score vector where each $\text{cred}(e_i)$ denotes the degree of credibility of edge $e_i$. Then, we can represent formula (1) as:

$$C = S^* \cdot C \hspace{1cm} (4)$$

Finally, we estimate the credibility scores of all text-image pairs by calculating the eigenvalues of matrix $S^*$.

The definition of support degree varies according to the type of target Web information that is in the form of text-image pairs and the situation in which the target Web information is used. Once the degree of support between two text-image pairs is defined for a given TIP graph, $G$, we can compute the degree of credibility for each edge in $G$.

### 2.3 Support based on typicality/peculiarity

Here, we introduce two basic aspects we used to analyze the support relationship between text-image pairs.

**Edge typicality.**

The first type of support is the *typicality* of node-to-node association of a target edge. The degree of edge typicality is the degree to which one node of the target edge is typical for the other node of the target pair when considering other edges. Intuitively, this degree can be estimated by checking how many similar edges the target edge has in a TIP graph.

For example, we can introduce the aspect of edge typicality to the case in Figure 3, where we wish to check the credibility of the target image from the viewpoint of the typicality of the image for the image label. Then, in given images that have the same product name (image label) for a target text-image pair as text-image pairs, the more images of a given product have a greater similarity to the target, the more credible it is.

Here, edge $e$ can gain greater support from edge $e'$ in Figure 4(a), if the node on the data type $X$ of $e$ ($e'[X]$) is more similar to that of $e'$ ($e'[X]$) and the node on the data type $Y$ of $e$ ($e[Y]$) is more similar to that of $e'$ ($e'[Y]$). Now, let $\text{sim}_X$ and $\text{sim}_Y$ denote the similarity between $e[X]$ and $e'[X]$ and the similarity between $e[Y]$ and $e'[Y]$. In Figure 3, $\text{sim}_x$ is the similarity between text labels (product name), and $\text{sim}_y$ is the similarity between images. There are actually two kinds of directions for node-to-node associations, i.e., $X \rightarrow Y$ and $Y \rightarrow X$. For instance, we can intuitively define the degree of support to estimate *edge typicality* by considering the direction of node-to-node association as:

$$\text{sup}_{X \rightarrow Y}(e, e') = \begin{cases} 
\text{sim}_X(e[X], e'[X]) \cdot \text{sim}_Y(e[Y], e'[Y]) & \text{if } \text{sim}_X(e[X], e'[X]) > \theta \\
0 & \text{otherwise}
\end{cases} \hspace{1cm} (5)$$

We can regard the case in Figure 3 as being that to estimate the typicality of text-image associations as the credibility of text-image pairs. However, suppose that we wish...
to check whether a text label (person’s name) for an image is valid as the credibility of a text-image pair as in Figure 1 (Figure 5 represents the TIP graph of the case in Figure 1). In this case, the credibility of the target text-image pair can be estimated by considering sup\(_{Y \rightarrow X}(e, e')\) to be the definition for the degree of support.

**Edge peculiarity.**

The second aspect is the peculiarity of the node-to-node association of a target edge among related edges. The peculiarity of a node-to-node association is the degree to which one node of the target edge is peculiar, unique, or unpredictable for the other node of the target pair in taking other edges into account. This intuitively means the degree to which similar data of a target edge are connected by different data of the target edge in a TIP graph. For example, the peculiarity of node-to-node association can be introduced to the case in Figure 2. In this example, suppose that a Web page insists that Hamburger H in an image of the Web page is peculiar compared to other hamburgers and Hamburger H page B. A recent widely accepted interpretation is that images do not always correspond to their text labels.

One of the most serious issues in the credibility of images is that images do not always correspond to their text labels. For example, fake photos of a certain product are often on display with the product name on auction sites. We focus on two practical cases in this section to discuss the credibility of images: the (1) consistency between image and text and (2) image peculiarity for text. Then, we propose practical methods of analyzing the credibility of images by using the model described in Section 2.

### 3. Analysis of Credibility of Image

There can be various types of text-image credibility given images and their texts such as labels and text captions. For example, there are portraits of a certain Japanese historical personage on two Web pages in Figure 1. Although they illustrate the same person, the image captions indicate two different names: Ashikaga Takauji on page A and Kohn Moronao on page B. A recent widely accepted interpretation is that the person in these portraits is Kohno Moronao, and not Ashikaga Takauji.

To check which text label corresponds the most to the image in Figure 1, one possible way is to collect similar images from the Web and to check which text label is more frequently attached to similar images in the TIP graph in Figure 5. Let us suppose that we estimate text-image pairs that meet this condition in our model. Here, X denotes a set of historical people’s names, Y denotes a set of images for their names, and E denotes a set of edges between texts and images. Then, we can estimate the credibility of text-image pairs by regarding sup\(_{Y \rightarrow X}\) as the definition for the degree of support.

Even if one text label corresponds more to an image than any other labels, if the image is not typical for the text label, it is difficult to determine that there is consistency between the image and text label. Therefore, we also have to simultaneously consider sup\(_{X \rightarrow Y}\) as another degree of support to estimate the credibility of text-image pairs.

Finally, credibility scores C of text-image pairs E is calculated by:

\[
C = \alpha \cdot S_{Y \rightarrow X} \cdot C + (1 - \alpha) \cdot S_{X \rightarrow Y}
\]  

\[
C_{X \rightarrow Y} = S_{X \rightarrow Y} \cdot C_{X \rightarrow Y}
\]
Here, $S_{Y \rightarrow X}$ is a matrix for the degree of support that is defined by $S_{X \rightarrow Y}$, and $C_{X \rightarrow Y}$ is a vector of credibility scores that are calculated with the matrix for the degree of support, $S_{X \rightarrow Y}$. The $\sup_{X \rightarrow Y}(e, e')$ between edge $e$ and $e' \in E$ is calculated with $\sim_Y(e[X], e'[X])$ and $\sim_Y(e[Y], e'[Y])$.

We will discuss our approach to dealing with this using our model below. We selected name pairs of historical people in images as one example and devised a practical method of evaluating the credibility of such pairs from the viewpoint of consistency.

3.1.1 Practical implementation

When we estimate the credibility of name pairs of historical people in images in our model, the most important thing, in practice, is how to calculate similarity in people’s images as one example and devised a practical method of related name pairs of historical people in images.

In this implementation, we define $\sim_Y(e[X], e'[X])$ as:

$$\sim_Y(e[X], e'[X]) = \begin{cases} 1 & \text{if } e[X] = e'[X] \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

The $\sim_Y(e[Y], e'[Y])$ is defined as similarity between images $e[Y]$ and $e'[Y]$.

The calculation of image similarity and collection of related text-image pairs to analyze credibility are explained below.

Extraction of visual features and calculation of similarity.

There are numerous methods of extracting visual features from images. Local descriptors are particularly attractive because they can be used to extract excellent visual features, which are invariant to image scaling or rotation. We used SIFT local descriptor [12] to calculate visual similarity. By applying SIFT to an image, we can extract key points that represent the image’s local features. We define the image similarity between images $i_1$ and $i_2$ as the number of key points shared between $i_1$ and $i_2$ divided by the mean number of key points extracted from $i_1$ and $i_2$.

Collection of name pairs of historical people in images.

If the person’s name attached to an image is incorrect, we would expect other people’s names to appear around similar images on the Web. We therefore collected images by using the person’s name of a target image-entity pair and a Web-image search engine. We extracted people’s names from the texts surrounding the images we collected by using a previously prepared dictionary of 7225 people’s names that often appear in Japanese history. After collecting their names, we input each person’s name as a query to a Web-image search engine and collected the top $N$ results. Pairs of images that we obtained and the name that was used were considered to be related IH pairs.

The practical implementation to analyze the credibility of analysis for name pairs of historical people in images (IH pairs) involved six steps in the system workflow.

1. Receive a target IH pair as system input.

2. Collect images by using a Web image search engine and the name of a historical person in the target pair.

3. Extract other names of historical people from surrounding texts of collected images by using a dictionary of people’s names.

4. Collect images by using the collected names of historical people.

5. Extract visual features from the collected images and those of the target to calculate the similarity in images.

6. Calculate the credibility score for the target IH pair by applying formula 11 to the collected text-image pairs and the target.

3.2 Image peculiarity for text

One of the most critical cases for establishing text-image credibility is where a description is overly exaggerated for an image. Figure 5 has one example, where the text on a Web page insists that Hamburger H is different to any other hamburgers and such a hamburger has never been displayed. If that is not a fact, it can mislead people who visit the Web page.

One possible solution to detecting such exaggerations is to collect images of a target product and similar products and compare a target image with ones that are collected. If the target product is actually more typical than any other product (the text description does not exaggerate the image), there should be few similar images in the set of collected images.

We can estimate the credibility of text-image pairs that meet this condition in our model by regarding $\sup_{X \rightarrow Y}$ as the definition for the degree of support. Then, credibility scores $C$ of text-image pairs are calculated by:

$$C = S_{X \rightarrow Y} \cdot C \quad (11)$$

3.2.1 Practical implementation

We defined $\sim_Y(e[X], e'[X])$ by the normalized edit distance between product names $e[X]$ and $e'[X]$ in the practical calculation of $\sup_{X \rightarrow Y}(e, e')$ between edge $e$ and $e' \in E$. Image similarity $\sim_Y(e[Y], e'[Y])$ is defined by using the same definition as in Section 3.1.1. We focused on the credibility of images of hamburger products in this implementation, and we prepared a list of the names of hamburger products in advance.

The practical implementation to analyze credibility involved four steps in the system workflow.

1. Receive a target “product name (hamburger name)”-“product image” pair as system input.

2. Collect images for the target product and images for all products in the hamburger list by using a Web image search engine.

3. Extract visual features from the collected images and the target to calculate image similarity.

4. Calculate the credibility score of the target pair by applying formula 11 to the collected text-image pairs and the target.

4. APPLICATION AND CASE STUDY
4.1 ImageAlert

One necessity for ensuring credibility is to support users who assess the credibility of suspicious text-image pairs. We propose ImageAlert, which is a prototype application for checking the credibility of text-image pairs when browsing the Web. We have been developing a version of ImageAlert for Mozilla’s Firefox. The current ImageAlert enables users to check the credibility of “historical person’s image”-“historical person’s name” pairs and “hamburger’s image”-“hamburger product’s name” pairs. The two kinds of credibility are estimated with the method proposed in Sections 3.1 and 3.2. Figure 6 outlines an example of use, where the user doubts whether an image on a Web page is that of Ashikaga Takauji or not. When an image-object name pair is input into ImageAlert, the system checks other possible people’s names for the target image. Moreover, it returns more credible images for the person’s name that is input. These functions support users in finding whether the target text-image pair is credible or not. In addition, it enables them to identify more credible text-image pairs.

4.2 Case study

Here, we will explain various case studies in which we used our proposed application of ImageAlert, which demonstrate the effectiveness of our approach in detail.

We describe the credibility of three “historical person name”-“person image” pairs and one “hamburger product name”-“hamburger image” pair that were on some real Web pages in these case studies. If they were not sufficiently credible, we expected that ImageAlert would give us more credible images for object names and object names that conflicted on images similar to the target image.

Figure 7 shows the results where ImageAlert checked a portrait on a Web page about Saigo Takamori. We manually judged the portrait to be a poor copy and assessed its credibility to be an image of him by using ImageAlert. According to Figure 7, the system evaluated that the credibility of the target image-name pair was much lower than the most credible image (44.6 < 73.6). This is because the target image did not have similar images of Saigo Takamori on the Web. More credible images that the system output were actually correct portraits that people generally agreed were his portraits.

Figure 8 has the results for where ImageAlert checked a portrait on a Wikipedia article about Ashikaga Yoshimitsu. We manually determined that the portrait was real and assessed the credibility of the image of him by using ImageAlert. According to Figure 8, the system determined that the target image-name pair was more credible (70.5) than any other images of him. This is because the target image had similar images of him on the Web.

The system succeeded in measuring the credibility of the target image-person name pairs for the examples in Figures 7 and 8. However, it failed to assess the credibility of the target image-person name pair in Figure 9, where we selected an image that represented Ashikaga Takauji as input. According to recent historical studies, this image has been found to be a portrait of Kohno Moronao, and not Ashikaga Takauji. Therefore, we expected that ImageAlert could determine the target image was not credible as an image of Ashikaga Takauji, but the analysis failed. When we checked the reason for this failure, we found that when we searched for images of Ashikaga Takauji by using Web image search engines, we obtained a lot of images similar to the target image for Ashikaga Takauji. On the other hand, we also ob-
There have been some researchers who have estimated the credibility of a “specific” type of information and supported judgments on it. Vuong et al. proposed a method of finding controversial information in Wikipedia articles [18]. Kittur et al. described a tool for characterizing conflicts between sentences in Wikipedia articles [9] and evaluated what effects it had on users’ perception of the trustworthiness of Wikipedia articles [10]. These studies were based on editing histories and they were focused on content bias as one factor in trustworthiness.

In the field of multimedia research, some recent studies have focused on image or video forensics as one issue concerning the credibility of multimedia content. The aim of these studies has mainly been to detect intentional manipulation and copying of images to deceive users. Gloe et al. discussed the reliability of techniques for tracing the origins of digital camera images [7]. Zhu et al. proposed a method of retrieving duplicated images in different videos [21]. Even if information publishers do not manipulate images to deceive users, the content of image information often does not correspond to its surrounding text. Techniques of detecting copies are ineffective in such cases. Therefore, methods of semantically checking the credibility of content such as that with our method are also often necessary.

There have been some studies focusing on helping users judge credibility. Ennals developed Dispute Finder, the system that searches the Web for counter sentences when users judge credibility. Yamamoto et al. developed a system to aggregate Web information to judge the credibility of doubtful claims, such as comparative claims, temporal change of typicality of the claims, and people’s sentiment for claims on the Web [20]. For support systems to judge the quality of Videos, Nicholas et al. proposed a tool to analyze, collect, and share criticisms for the videos, especially about politics [2].

6. CONCLUSION

This paper presented a generic model for estimating the credibility of text-image pairs by using the idea of “support relation” between pairs. Our model was specifically focused on edge typicality and edge peculiarity as basic degrees of support between text-image pairs. We also implemented a prototype system called ImageAlert, which supports users in determining whether text-image pairs are credible.

There are several unresolved issues in enhancing the analysis of credibility with our model. We need to brush it up and conduct large-scale experiments to evaluate it. We also need to discuss a method of automatically determining optimal degrees of support to enable the credibility of text-image pairs.
pairs to be analyzed with our model. This should be automatically determined for users in practice. We believe that such work would enhance our analysis of credibility for Web information.

7. ACKNOWLEDGMENTS

This work was supported in part by a MEXT Global COE Program “Informatics Education and Research Center for Knowledge-Circulating Society” (Project Leader: Katsumi Tanaka), a MEXT Grant-in-Aid for Scientific Research on Priority Areas “Cyber Infrastructure for the Information-explosion Era”, Planning Research “Contents Fusion and Seamless Search for Information Explosion” (Project Leader: Yusuke Yamamoto, Grant#: 211243), and the National Institute of Information and Communications Technology, Japan.

8. REFERENCES