How Do Humans Access the Credibility of Weblogs: Qualifying and Verifying Human Factors with Machine Learning

Yonggeol Jo
Ajou University
Suwon, Republic of Korea
yonggeol93@ajou.ac.kr

Minwoo Kim
Ajou University
Suwon, Republic of Korea
rlaalsdn4242@ajou.ac.kr

Kyungsik Han*
Ajou University
Suwon, Republic of Korea
kyungsikhan@ajou.ac.kr

ABSTRACT

The purpose of this paper is to understand the factors involved when a human judges the credibility of information and to develop a classification model for weblogs, a primary source of information for many people. Considering both computational and human-centered approaches, we conducted a user study designed to consider two cognitive procedures–(1) visceral, behavioral and (2) reflective assessments–in the evaluation of information credibility. The results of the 80-participant study highlight that human cognitive processing varies according to an individual's purpose and that humans consider the structures and styles of content in their reflective assessments. We experimentally proved these findings through the development and analysis of classification models using 16,304 real blog posts written by 2,944 bloggers. Our models yield greater accuracy and efficiency than the models with well-known best features identified in prior research.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**;

KEYWORDS

Information credibility; Weblogs; Social networking services; Cognitive machine learning

*Corresponding author

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300904

ACM Reference format:
Yonggeol Jo, Minwoo Kim, and Kyungsik Han. 2019. How Do Humans Access the Credibility of Weblogs: Qualifying and Verifying Human Factors with Machine Learning. In *Proceedings of CHI Conference on Human Factors in Computing Systems Proceedings, Glasgow, Scotland UK, May 4–9, 2019 (CHI 2019),* 12 pages.
https://doi.org/10.1145/3290605.3300904

1 INTRODUCTION

It is commonplace for people to share and look for information through social networking services (SNSs) [23]. Unlike the sources that provide unilaterally objective (or formally written) information through traditional online channels (e.g., news, public institutions), the SNS has become a channel which provides not only both formal and informal information (e.g., hobbies and interests) but also an environment in which people can interact through commenting or liking.

In particular, weblogs, one of the SNSs, show differences in terms of users and information provision when compared to other media. First, according to statistics, blogs are used as a key source of information in Asia and Latin America [31]. Second, anyone (e.g., both professionals and ordinary people) anonymously can write blog content, sharing their own experiences and in-depth information [24]. Blogs are free from the restrictions of formats (e.g., number of letters, length, font; whereas Twitter has a 140-word limit, and Facebook users tend to write short messages) and can use various types of metadata (e.g., video, image, map). Third, readers can be aware of other readers' opinions through additional features or activities (e.g., comment, sympathy, link, scrap). This can be seen from the fact that the ratio of product and brand reviews in blog usage is high, because such reviews need longer text (detailed description) and more metadata [9, 31].

As the reliance on the acquisition of information increases, the problem of the credibility of information is also emerging. Readers cannot easily judge the credibility of information, since there is no easy way to confirm whether the presented information is credible. This leads to a situation where the credibility of information is often judged by heuristic cues, along with the background of the user (e.g., knowledge, experience, and expertise). This makes rational judgment more

Type	Best feature identified	Castillo et al. [8]	Mukgerjee et al. [33]	Lu et al. [30]	Han [19]	Benevenuto et al. [3]	Li et al. [29]	Lee et al. [27]	Ferrara et al. [13]
Content	URLs (count, ratio, existence)	✓			✓	✓			
	Questions (count, ratio, existence)	✓			✓				
	1st person pronouns (count, ratio, existence)	✓					✓		
	2nd person pronouns (count, ratio, existence)			✓			✓		
	Number of hash tags				✓	✓			
	Length of text		✓						
	Content similarity		✓	✓	✓		✓		
Activity	Author profile length	✓			✓		✓		
	Has an author's profile image							✓	
	Account age	✓				✓			✓
	Number of followers	✓			✓	✓			
	Number of posts						✓		
	Number of posts the user replied to (fraction)					✓			
	The number of posts per day (time interval)		✓	✓				✓	✓
Sentiment	Sentiment score	✓			✓				✓
	Sentiment positive words	✓			✓		✓		
	Sentiment negative words	✓			✓		✓		
	Neutrality of information						✓		
	Emoticon smile, frown (emogji)	✓							

Table 1: Summary of best features identified in computer-based approach.

difficult due to the limitations of the human brain with respect to information perception and processing [10, 38]. In fact, we have witnessed the creation of a great deal of fake information that has led to many personal and social issues (e.g., 2016 US Presidential Election[1], 2017 Angela Merkel selfies with terrorists[2], 2017 French presidential election)[3].

To solve this problem, research has proposed methods to measure and predict the credibility of online information through computational analyses or classification modeling [3, 8, 13, 19, 27, 29, 30, 33, 40]. However, such efforts are somewhat disconnected from the theoretical understanding of the cognitive processing involved in the decision of information credibility. Many theoretical studies [11, 12, 22, 34, 37] have argued that a reader's decision making is influenced by two types of thinking, namely (1) visceral & behavioral and (2) reflective; yet, little research that employs a computational modeling approach to the understanding of information credibility has applied such thinking in the labeling process and studied how the features identified from such thinking influence model performance.

Since blogs are often top search engine results, more blogs are being created and readers' dependence on the information from blogs is increasing. The ability to freely express bloggers' genuine experiences, thoughts and feelings on specific topics of interest means that blogs are often seen by ordinary people as sources of honest and credible information [9, 20]. Surprisingly, however, it has been reported that a growing number of people are suffering mentally and financially because of non-credible information on blogs [21, 42]. Examples include fake images or reviews of hotels, restaurants, or products posted by a user who never had corresponding experiences (mostly having been asked by an advertisement company to write fake reviews). This is because blogs can provide both formal and informal content with a variety of visual tools in a more flexible fashion than other types of SNSs, making people believe that a blog post is more reliable compared to posts on other SNSs. This difference suggests that the understanding of blog information credibility should be considered and examined.

In this paper, we aim to answer the following research question:

"What is a reader's cognitive processing of information credibility in blogs and how does the understanding of such processing influence a computational analysis of information credibility in blogs?"

2 RELATED WORK

Studies of information credibility have been conducted for a long time. In discussing related work, we focus on two main approaches to understanding and examining credibility in the existing research: (1) discovering effective features and developing a computational model, and (2) finding important factors that influence a user's perception of credibility on the web through user studies.

[1] https://www.npr.org/2018/04/11/601323233/
[2] https://www.bbc.com/news/world-europe-38599385
[3] https://www.bbc.com/news/world-europe-39495635

Paper	Factor	Survey Question (Select all criteria which you considered to evaluate credibility)	Features for modeling
Burgoon et al. [6] Vissher et al. [41]	Qualified/unqualified	- Sincerity of the post	- The number of grammar errors - Text/image count (effort text/image ratios)
Fogg et al. [17]	Visual design	- Visual media presence or absence - Blog appearance (cover image, design) - The presence of emotion stickers (emojis) - Text style (font, font size, bold)	- The number of media - The number of stickers - Presence of map - Font type conversion - Font size conversion - Bold text length ratio - Color text length ratio
	Information design/structure	Coherency of the post (alignment, structure) - Information design/structure	- Alignment - Structure
	Company product link	- The post contains product or service URL	- Existence of URL
	Bias of information	- Neutrality of information	- Negative and positive sentiments - Subjectivity - Polarity - Sentiment differences
Lewandowsky et al. [28]	Coherency of message	- Coherency of the post (structure, text, image)	- Text/image count (effort text/image ratios)
	General acceptability: do others believe this information?	- The number of comments, hearts (favorites, likes)	
	Timeliness of blog content (frequent updates of content)	- Difference of post interval compared to previous posts	
Banks [1] Burgoon et al. [6] Fogg et al. [17] Visscher et al. [41]	Experienced/Inexperienced	- Author has an experience (The post was, or was not, written by an author who had experience)	- Existence of personal pronouns (first/second person ratios)
	Authentic (exclusive coverage of an interesting topic)		
	Expert/inexpert	- Author expertise (The author is, or is not, an expert in this area)	
	Professional/Unprofessional		
	Information focus (having simple words use, breath of information available, specific domain)	- Total post of this author - The use of easy or professional vocabulary	
Weil [44] Morris et al. [32]	Profile information	- Author information (name, profile)	
	Focused (delving into a specific time; establishing a niche of personal passion)	- Uploaded time	

Table 2: Summary of factors affecting people's decision-making, as found in a human-based approach. We converted the factors into survey questions and investigated which factors would be selected by humans (i.e., blog readers) as the criteria (converted features) of the credibility of blog posts. Some of the features were found in multiple literature. Features with blank cell were not used in modeling.

Computer-based approach

Computational approaches have been used to construct models that classify credible and non-credible information by using machine learning with various features (e.g., content, user's information, network). For data collection and annotation, researchers have used various methods. For example, researchers have used crowd-workers [35] or asked study participants to annotate the credibility of the content, defined data based on the characteristics of existing spam or fake text, or used data already labeled by the system (such as Yelp data [33]). The researchers then identified a set of features that perform well in distinguishing credible and non-credible information. For example, Carlos et al. [8] divided the features used to evaluate credibility into four categories according to their scope (i.e., message, user, topic, and propagation), then evaluated the credibility of newsworthy tweets and found the 15 best features. Their results indicated that the negative sentiment term and the inclusion of a URL occurred more often in the credible news tweets. For non-credible news, the positive sentiment term appeared more. Similarly, Mukherjee et al. [33] used Yelp data, found five strong features (i.e., maximum number of reviews, percentage of positive reviews, review length, reviewer deviation, and maximum content similarity) and presented a model with an accuracy of 84.1%.

Moreover, Li et al. [29] and Lu et al. [30] showed that the rate of personal pronouns in the sentence, the similarity of text, and the number of articles previously written by users are important features for finding fake content. Through a user study, Han [19] asked people to read multiple tweets from the same poster and evaluate their credibility, then built a model with an accuracy of 93% using author profile, syntax, content similarity, sentiment, and linguistic features. Benevenuto et al. [3] constructed a model using 39 content and 23 user behavior attributes, and found the top 10 features used in the model: fraction of tweets with URL, age of user account, URLs per tweet, fraction and number of the followers per followees, the fraction and the number of tweets to which the user had replied, and the average number of hashtags per tweet. Lee et al. [27] proposed a social honeypot system to detect social spammers and found account age, URL per tweet and content similarity were the best features for spammer detection. Ferrara et al. [13] and Stringhini et al. [40] classified spammers and non-spammers using user profile information and behavior features.

In summary, building a model with feature selection and presenting good performance is the main goal of a computer-based approach. The best features mentioned in prior studies are in Table 1.

Human-based approach

The human-centered approach combines the theoretical element of cognitive science with an empirical method that uses surveys and interviews. Many studies have been conducted to understand the factors affecting cognitive processes and evaluations. Table 2 summarizes such factors. We will discuss features for modeling in the later sections.

Theoretical understandings. In the cognitive process, users look for cues of deception or misinformation to evaluate the credibility of the information. [28] However, in many real cases, it is unlikely that people spend enough time and effort evaluating information credibility. Fogg et al. [17] found that users' credibility determinations were significantly influenced by visual factors rather than content and source information. This shows that the user does not consider all factors when evaluating credibility but rather evaluates only the specific components [16]. This is because a person has a limited cognitive capacity and does not want to think too much. Similarly, according to the Bounded Rationality theory of cognitive science [38], people are known to have good performance when considering satisficing instead of the optimal decision, due to limited time, dynamic conditions and knowledge [25].

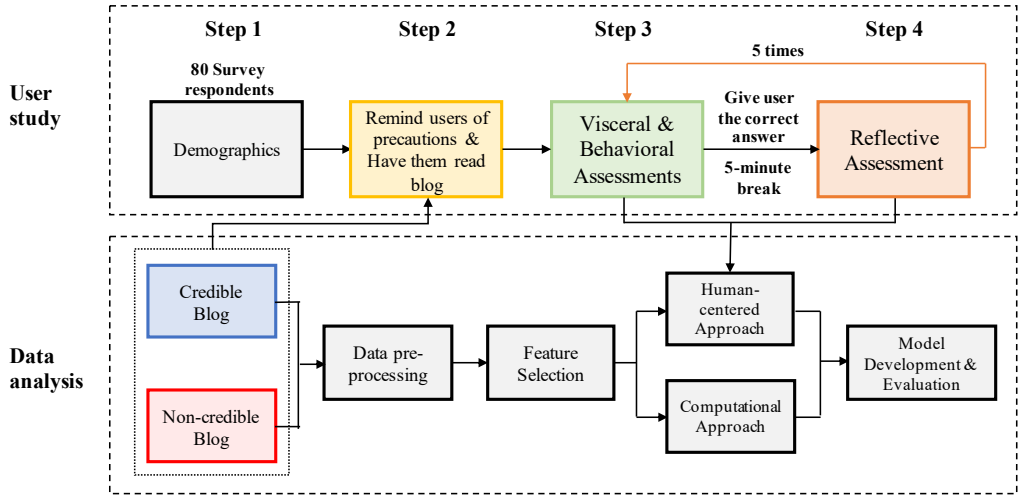
Such people's cognitive behaviors have been studied in psychology and cognitive science with respect to dual-process and dual-system theories [11, 12, 22, 37]. Kahneman and Frederick [22] use the terms "System 1" and "System 2" for decision-making. Evans [11], Stanovich [12], and Samuels [37] used "Type 1" and "Type 2" processing. Although the terminologies of the human cognitive process vary slightly among researchers, they mostly agree that there are two types of thinking: *intuitive thinking* and *reflective thinking*. On the one hand, intuitive thinking is fast and instinctual. Because it happens quickly and automatically, decisions and tasks associated with it feel easy and natural. However, it is also prone to perception errors, bias and other experiential influencers. On the other hand, reflective thinking is slow and analytical, requiring time and effort. It requires focused mental activity, and decisions and tasks associated with it feel complex and demanding. However, it leads to more reliable and careful decisions than intuitive thinking.

In the Human-Computer Interaction field, Norman [34] similarly defined such processes in three levels, where the first two (visceral and behavioral) match intuitive thinking and the last (reflective) matches reflective thinking. The visceral level is responsible for the ingrained, automatic and almost animalistic qualities of human emotion, which are almost entirely out of conscious control. The behavioral level refers to the uncontrolled aspects of human action, where we unconsciously analyze a situation so as to develop the goal-directed strategies most likely to prove effective in the shortest time, or with the fewest actions possible. Lastly, the reflective level is where deep understanding develops, where reasoning and conscious decision-making take place. To summarize, while the visceral and behavioral levels are subconscious and as a result respond rapidly, reflection is cognitive, deep, and slow, and often occurs after events have happened.

Empirical understandings. Many HCI studies have aimed to find factors that affect people's credibility evaluation. Unlike the computational approach, a survey or interview method is widely used. For example, Yang and Lim [45] found that users tend to trust institutions according to their interactivity level. Burgoon et al. [6] and Visscher et al. [41] attempted to measure credibility using relational communication (i.e., trustworthy, expert, reliable, intelligent, professional, experienced), but did not achieve significant results. Banks [1] interviewed 30 active SNS posters and found that credible blogs are focused (i.e., delving into a specific time, establishing a niche of personal passion), authentic (i.e., having exclusive coverage of an interesting topic), and insightful (i.e., offering in-depth opinions and rich personal experience). Banks [1] and Weil [44] suggested that the timeliness of the content is a key feature of its credibility. Fogg et al. [17] evaluated the reliability of two websites and identified credibility factors, including visual design, information design/structure, information focus, and company motive. Based on cognitive psychology, Lewandowsky et al. [28] summarized the mechanisms of human credibility assessment as consistency of message, coherency of message (how well a piece of information fits a broader story that lends sense and coherence to its individual elements), credibility of source, and general acceptability. Morris et al. [32] found that not only the elements (e.g., grammar/punctuation, hashtags, URL) of posts but also information about the author (e.g., follower, personal image, location, posting dates) affected the user's credibility evaluations.

Limitation and opportunities

Based on previous research results, we found the following limitations in the study of information credibility. The first relates to the labeling process of information credibility. In the computational approach, data collection is often dependent on human annotation. The criteria for evaluating credibility differs from person to person, which however has less considered or neglected in many computational studies. Second, the human-centered approach has contributed to understanding humans' cognitive processes by grasping factors affecting their credibility judgment. However, since a person does not judge all elements as having the same weight when assessing the reliability of an SNS, there are relatively few studies of the factors considered in evaluation



- Verifying effective features derived from the previous work and user study
- Developing model to classify whether information is credible or non-credible

Figure 1: Overall process of our study.

and verification. Thus, our study strives to address these limitations of previous studies (that the computational and human-centered approaches are somewhat disconnected) of information credibility by conducting a study to understand the cognitive process of blogs, and more broadly of SNSs.

3 OUR DEFINITION OF CREDIBILITY

Although many studies have been done on credibility, its definition and the factors used to evaluate it vary somewhat between researchers. In previous studies, the factors involved in determining information credibility include information sources, information semantics (e.g., accuracy or neutrality), the appearance characteristics of information (e.g., design and editing status of the website), and the sources of information. Fogg and Tseng [15] suggested that trustworthiness and expertise are key components that most researchers commonly consider in the study of credibility. Trustworthiness is the belief that the information provided is the honest opinion of the information source and not distorted; its components include morality and good intention. Expertise is the belief that the information source has the knowledge and ability to make plausible statements about the subject, and is generally measured by attributes such as the possession of expertise, richness of experience, and depth of content.

Based on the definitions of credibility in previous studies, we regard "information containing the genuine opinion of a knowledgeable user in the field" as credible information, and on the contrary, "information from a user who does not have experience of the specific field or expert knowledge" as non-credible information. Given such definitions, in this paper, we propose a new method for classifying credible and non-credible information and for building a credibility classifier based on various characteristics of blog content.

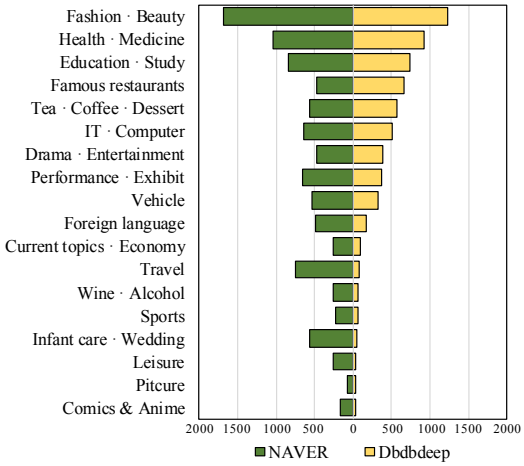


Figure 2: Distributions of blog post categories from the two sources.

4 USER STUDY

Data definition and collection

The overall process of our research is shown in Figure 1. We chose NAVER[4], the most widely used web portal (75.2% Korean use)[5] in Korea; NAVER blogs dominate 67.1% of total blog usage in Korea[6]. In the following section, we explain how we chose credible and non-credible sources.

Credible and non-credible sources

By reading a blog post, it is difficult to judge whether the post was written by a credible or non-credible source. Fortunately, NAVER reviews blogs from more than 30 categories (e.g.,

[4]https://www.naver.com/
[5]https://news.joins.com/article/22435278
[6]http://www.blogchart.co.kr

Step 1. Demographics & User behavior
1-1. Demographic Information
• Age
• Blog management experience & period
1-2. User activity on weblogs
• Purpose of blog use
• Type of information on blogs
1-3. Blog search behavior (5-point Likert scale)
• I only consider the content to evaluate credibility
• I consider other blog posts written by the same author to evaluate credibility
• I consider both a piece of content and author information to evaluate credibility
Step 2. Precautions
2-1. Remind of the purposes of this survey (blog credibility evaluation) and then introduce a study procedure
• Read the blog post and answer given questions
• It takes about 5 minutes to read each blog
• This process is repeated 6 times
Step 3. Visceral & Behavioral Assessment & Feature selection
3-1. Visceral & Behavioral credibility assessment
3-2. Factors considered when assessing the credibility of blog
Step 4. Reflective Assessment & Feature selection
4-1. Reflective credibility assessment
4-2. Factors considered when assessing the credibility of blog after knowing the correct answer

Table 3: Design of the survey.

language, foreign language, fashion, beauty, domestic travel) every month based on five guidelines (i.e., experience, trust, commerciality, copyright, activity) and selects blogs that provide excellent and reliable information in the field, which are named as "power bloggers." Therefore, we define a power blogger as a credible source. In the case of a non-credible source, Dbdbdeep[7] is a website that recruits people who do not have a real experience to write reviews for the customers of the website. Various types of items are reviewed, such as restaurants, hotels, hair shops, cosmetics, etc. Review posts are created based on pre-defined guidelines, descriptions, and images or videos provided by the customers. Such blogs are written by people who have no experience or knowledge in the field, and many of the posts are written for advertising.

Data collection

We first accessed the power bloggers' posts selected from NAVER and collected articles from the selected categories. In the case of Dbdbdeep, we crawled all blog posts available on the website. The crawling code was written in Python, and the data collection period was from April to July 2018. As a result, we collected 9,942 posts from 989 users on NAVER and 6,362 posts from 1,955 users on Dbdbdeep. Both sources have the same list of 18 categories as shown in Figure 2.

Study design

The user study has three goals:

- Goal 1: Identify factors affecting a user's blog credibility judgment and derive new features for modeling.
- Goal 2: Examine behavioral differences between visceral & behavioral thinking and reflective thinking.
- Goal 3: Investigate design elements of credible blogs and factors for detecting non-credible blogs.

We used SurveyMonkey and followed its service regulations to collect response data. Our study was reviewed and approved by the university's internal IRB. Only people who consented our study (the informed consent was presented at the beginning of the survey) could participate in the study.

The structure of the survey is shown in Table 3. A survey questionnaire consists of the items that correspond to the features that can be obtained from the blog. For a human-centered approach, we inferred the factors that are known to affect the judgment of a human's credibility as shown in Table 2. A total of 6 blog posts (3 blog posts for each credible and non-credible group) were randomly selected and presented to survey respondents. The survey was designed to have visceral & behavioral assessment and reflective assessment, along with an assessment of the credibility of each blog post. The survey consists of 4 steps:

- Step 1: Respondents were asked about their demographics and the purpose of their blog use. We rated their information-seeking behavior on blogs.
- Step 2: Respondents were asked to read the blog by accessing the blog URL. A total of 6 blogs (3 credible and 3 non-credible blogs; one blog presented and each time) were randomly presented.
- Step 3: Respondents were asked to viscerally judge each blog post by selecting items from the list that they used and thought important for credibility assessment (Visceral & Behavioral assessment).
- Step 4: After taking a five-minute break, the respondents were informed of the answers for the blog posts that they had evaluated. Then they were again asked to carefully select the items that they thought important for credibility assessment. We gave the respondents enough time to think (Reflective assessment).

Reflective trigger

For the presentation of the answers before reflective assessment (Step 4), many prior studies had employed the same condition used as a trigger to induce reflective thinking [4, 5, 7]. Butterfield and Metcalfe [7] found that people easily correct erroneous responses to general information questions with high confidence, so long as the correct answer is given as feedback. Boud et al. [4] emphasized that the promotion of reflection is often associated with post-practice methods of experience capture. Thus, giving the correct answer to user allows the respondents to reflect on their previous answers

[7] http://dbdbdeep.co.kr/

Human Judgements				
Type	Positive	Negative	Total	Human error rate
NAVER	162	78	240	0.33
Dbdbdeep	57	183	240	0.24
Total	219	261	480	0.28

Table 4: Summary of evaluation results by the participants.

and to give an opportunity to re-evaluate the blog content and pick their selection criteria in the reflection phase.

Survey respondents and initial results

The survey took about 12 minutes (on average), apart from the 5-minute break time. The total number of the respondents was 80 (42 males and 38 female; 24 university staff and 56 students), with an average age of 21.3 (3 respondents were in their 10s, 64 in 20s, 12 in 30s, and 1 in 40s). All respondents were invited to the university laboratory and compensated with a $5 gift certificate.

Measuring human performance is important for several reasons. First, there are few studies that provide baseline models [35]. Second, measuring the performance of a person gives credibility to the collected data and justification for constructing a computational model. Table 4 shows the results of human judgment with a 28% error ratio. Although the error rate is high, compared to a previous study that showed a human error rate of 40% [35], our study showed better results. This is presumably because blogs have more factors than other SNSs, such as expression flexibility (e.g., no length limit, font size, color), diverse metadata (e.g., video, image, map), and additional features (e.g., comments, sympathy, link, scrap).

Table 5 shows the result of the visceral & behavioral assessment and reflective assessment questionnaires. It appears that the user evaluates the visual factors (e.g., media presence/absence, style of writing) as a higher priority when assessing credibility during the visceral & behavioral assessment phase than in the reflective assessment phase. This shows that, similar to Fogg's findings [17], people's perceptions of blog credibility are mainly influenced by visual factors. In the case of the reflective assessment, higher priorities were placed on the flow of the blog (i.e., coherency) than on visual factors, and two factors that had not been considered in the visceral & behavioral assessment phase (i.e., author information and total post of this author) appeared in the reflective assessment.

In summary, the results of our user experiment indicate that the *coherency of the post* is considered important after the participants became aware of whether the blog post they read were from credible or non-credible sources. Using four features of coherency (i.e., alignment, structure, effort text ratio, effort text ratio) identified in our survey study 2, we

Rank	Visceral & Behavioral Assessment	Reflective Assessment
1	Author has an experience	Author has an experience
2	Sincerity of the post	Coherency of the post
3	Visual media presence or absence	Author information
4	Coherency of the post	Author expertise
5	Author expertise	Sincerity of the post
6	Neutrality of information	Total post of this author
7	The post contains product or service URL	Visual media presence or absence
8	Text style	Neutrality of information
9	Blog appearance	Blog appearance

Table 5: The difference in top items chosen selection between (1) the visceral & behavioral assessment phase and (2) the reflective assessment phase in the survey study. The item in the blue cell indicates increased rank in the reflective assessment. The item in the green cells only appears in the corresponding assessment phase.

aimed to verify the influence of these features on classification performance by comparing the classification models with and without them.

Feature selection

The computational approach has to date considered user characteristics (e.g., user profile information, upload interval, content similarity, followers, number of posts). However, in this study, we try to construct a model based only on content features, excluding user-related features. The reasons for this are as follows. (1) For power bloggers, user activity level is generally higher than for general users because of the greater exposure of blogs to the public. (2) Due to the structure of a blog, it is difficult to check the user-activity status on one screen, and the rate of judging based only on the article is high. (3) Survey results from Table 3 (Step 1-3) show that more users read blogs only (Mean: 3.59, SD: 1.21) than considering an author information to evaluate credibility (Mean: 2.64, SD: 0.93) (t_{158} = 4.33; p<0.001). Therefore, applying content features to credibility classification modeling is likely to reflect the "trust level of individual blog posts." We believe this is a better approach, considering the real uses of blog posts by people [39].

Feature description

We used three primary groups–content, sentiment and style features–that affect the judgment of a reader's credibility by employing the features identified in previous studies of modeling and the ones identified in our user study. Table 6 shows the features and descriptions that were used in previous work and developed by us. Below are the features derived from our survey.

Scope	Feature	Description
Content	Title length	The length of title
	Existence of URL	If contains URL, the value is 1, otherwise 0
	Effort text ratio	The length of the text by the average length of all texts in the same category
	Effort image ratio	The number of images by the average number of all images in the same category
	The number of question marks	The number of question mark '?'
	First person ratio	The number of 1st person words by the total number of words in text
	Second person ratio	The number of 2nd person words by the total number of words in text
	The number of tags	The number of hash tags
	The number of stickers	The number of stickers similar to Likes
	The number of media	The number of media objects (e.g., video, audio)
	The number of grammar errors	The number of grammatical errors
	Presence of map	If post contains map, the value is 1, otherwise 0
Sentiment	Positive ratio	The number of positive words by the total number of words in text
	Negative ratio	The number of negative words by the total number of words in text
	Subjectivity	Proportion of sentiment to frequency of occurrence
	Polarity	Percentage of positive sentence references among total sentiment references
	Sentiment differences	The difference between pos and neg words by the total number of words in text
Style	Alignment	Alignment of image and post (left, center, right, both)
	Structure	Arrangement of text and images (e,g., text-image-image-text-image)
	Font type conversion	If user changes the font types (e.g., Times New Roman), the value is 1, otherwise 0
	Font size conversion	If user changes the font size, the value is 1, otherwise 0
	Bold text length ratio	Amount of bold text by the length of text
	Color text length ratio	Amount of colorful text by the length of text

Table 6: Blog content features used in modeling and analysis. Features with blue colors indicate the ones found in the reflective assessment phase of the user study. Note that author-related features were not considered in modeling.

Content features. Content features include title length, existence of URL, effort text ratio, effort image ratio, the number of question marks, first person ratio, second person ratio, the number of tags, stickers, media, and grammar errors, and presence of the map. Among these, effort text ratio and image ratio are aspects of the coherency of the post.

- *Effort text ratio* : There are many topics and domains in blogs; thus, we believe that features should be designed considering the characteristics of the domain (e.g., the length of the blog post, the number of pictures). We believe this describes the coherency of the blog with respect to its related with other blogs in the same category. Therefore, for normalization, we divided the text length of the article by the average text length of the category.
- *Effort image ratio* : With the same criteria we applied for text ratio, the image count of the article is divided by the average image count of that category.

Sentiment features. Sentiment analysis has been conducted in news and blogs, and it has been shown that the sentiments of the content influence the credibility of the content. Sentiment data consist of time series of favorable (positive) and unfavorable (negative) words. Let p and n denote the number of raw positive and negative references, which occur a total of N times in the corpus. Subjectivity and polarity are calculated as follows [2, 18]:

- *Subjectivity* indicates a proportion of sentiment to frequency of occurrence. It is measured by subjectivity = $(p + n) / N$
- *Polarity* is a percentage of positive sentence references among total sentiment references. It is measured by polarity = $(p - n) / (p + n)$

Style features. For style-related features, we used alignment, font type conversion, font size conversion, bold text length ratio, and color text length ratio.

- *Alignment* : We used 4 alignments of images and posts (i.e., left, center, right, both). Items were counted according to the sorted sentences or images and used as a feature.
- *Structure* : The arrangement of text and images is measured. All texts in the data were divided into various cases of 5-gram text and image sequences, and term frequency-inverse document frequency (TF-IDF) with 13 features was used to vectorize the order of the structure.
- *Font type conversion* : Bloggers can change fonts (e.g., Times New Roman, Arial). If a given blog post has font type conversion, we assigned the blog post to 1, otherwise 0.

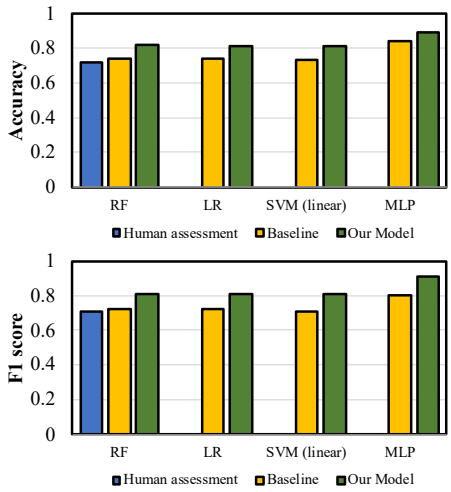


Figure 3: Performance of the different models. "Baseline" model is built with the well-known features identified in prior computational approach literature (Table 1, user features excluded). "Our Model" is built with the features that were founded from our user study. Our Model outperforms both the one with the existing important features and the human assessment.

- *Font size conversion* : Bloggers can change font size. If a given blog post has font size conversion, we assigned the blog post to 1, otherwise 0.
- *Bold text length ratio* : Bloggers can write some text in bold. We divided the length of the text that the user put in bold by the length of the entire text.
- *Color text length ratio* : Bloggers can also change the color of the text. The length of the text of which the user changed the color was divided by the length of the entire text.

5 RESULTS

Model development and performance

We finally investigated the performance of two models. The first model built as "baseline" with the features that were considered important in previous studies. The second model as "our model" is built with the features that were identified in our user study, which is the inclusion of the following style, neutrality of information factors and coherency components: structure, alignment, effort text ratio and effort image ratio.

We used the following machine learning algorithms for modeling: random forest (RF), logistic regression (LR), and linear support vector machine (SVM). We further used Multi-Layer Perceptron (MLP) neural network models with four hidden layers (using relu). The models predict whether a blog post is credible or not (binary classification). The MLP model was trained with 50 epochs and the ADAM optimization algorithm. For all models, we used 5-fold cross validation to avoid overfitting. We used the Keras (https://keras.io/) and Scikit-learn (http://scikit-learn.org/) libraries. The performance results of the models with four algorithms are shown in Figure 3. We can see that the MLP model outperformed other models by showing 89.2% of accuracy and 90.9% of F1-score. Overall, for credibility detection, our model based on MLP yielded the best performance, compared to the human assessment (Acc: 71.9%, F1: 70.6%) and the baseline (Acc: 84.2%, F1: 80.2%). This demonstrates the important role of the features derived from our user survey in model performance.

Rank	RF	LR	χ^2
1	Effort image ratio	The number of grammar errors	Existence of URL
2	Alignment	Effort text ratio	Structure
3	Existence of URL	Strucutre	Presence of map
4	The number of grammar errors	Alignment	The number of stickers
5	Structure	Effort image ratio	Font size conversion
6	Effort text ratio	Existence of URL	Font type conversion
7	The number of stickers	The number of stickers	The number of tags
8	Title length	Font size conversion	Polarity
9		First person ratio	Alignment
10		Negative ratio	The number of media

Table 7: Most important features, as identified in different modeling results. Increased ranked features in the reflective assessment were top-ranked.

Feature verification

Lastly, we examined the top features used in this modeling. This is to see how influential each feature is in model development. We obtained the importance of features from (1) logistic regression, (2) random forest, and (3) Chi-square [46]. Table 7 presents the top 10 most important attributes for each model (only 8 features were greater than 0.0 from the RF results). As a result, the structure, alignment and effort text/image ratios were found to be important in modeling. This means that the *coherency of the post is significantly associated with the credibility of the content.* We believe that four important components of coherency identified in our study should be highlighted and informed to readers for their reliable credibility assessment on weblogs.

6 DISCUSSION

Summary and design implications

The cognitive process of a human is composed of various factors such as abstraction, searching, learning, decision-making, inference, analysis, and synthesis [43]. This process is known to make high-dimensional thinking. Sometimes, it produces errors, misunderstandings, or wrong decisions.

Because of human errors caused by the complexity of such process (e.g., the limited capacity of conscious thinking, the tendency to protect one's feeling of competency, the weight of the actual problem, and forgetting) [10], understanding of such social phenomenon only through large-scale data and computational approaches may not offer suitable results or insights that are applicable to other domains.

The contributions of our work are twofold. First, we presented a comprehensive analysis on understanding the cognitive factors involved when a human judged information credibility. We examined the factors of a user's cognitive process through the comparison of the features identified in our study with those identified in previous studies. This was done through data- and model-driven analyses, which not only differed from many prior studies with psychological approaches [26] but also substantiated these studies in such a way that different factors of information credibility perceptions occurred in visceral & behavioral and reflective assessments.

Second, we developed a classification model for weblogs, in terms of the presentation and composition of the information. If human factors are successfully understood and applied to the computational approach (e.g., machine learning), we may achieve more meaningful results. Our findings of four representative coherency features–structure, alignment, effort image/text ratios–show that such cognitive process is efficiently measured. This is because those features are not only easily calculated and vectorized but also demonstrate their strong influence on model performance compared with existing features such as those from natural language processing.

Some of the elements studied in our work give design opportunities. For example, the existence of a product or service URL, and font style only appeared in the visceral & behavioral assessment phase. Since most people will end up with making a certain decision in this phase, highlighting the reliability of an URL link (a link can be a spam) would be useful for readers' evaluations of information credibility. In addition, since the coherency of the post is found to be highly important but tend to be overlooked by people during the visceral & behavioral assessment phase, the components associated with the coherency should be highlighted up-front to readers and some types of visual presentations would be helpful and useful (e.g., level of alignment of the post, level of the balance of the compositions of images and text, and the structural patterns of images and texts).

Limitations and future work

Although our study presents interesting insights, there are some limitations that will be addressed in our future work.

First, even though it has been reported that a person tends to believe and become more confident by visual attributes than printed words [36], our study did not sufficiently use image-related features but only the effort image ratio. Therefore, we will examine the influence and performance of features that are associated with images such as background or cover image of the blog page, profile image of a blogger, and image characteristics in the blog post. Various features can be extracted such as brightness, saturation, hue, colors, and objects in the image. In addition, we will apply important features found in our study to other social networking sites and verify the importance of these features.

Second, although we tried to have many categories of the blog posts to make our findings more comprehensive, our results may still be influenced by some topics with high frequency. Prior studies indicate that people's credibility judgment can vary depending on topics [14, 32], therefore, it would be necessary to look into the model performance with the same set of the features found in this study by category.

Third, as Table 4 shows, 72% of the answers were correct for detecting credible or non-credible blog posts. Here, our preliminary analysis of the one-fourth of the wrong answers indicated that, among the many factors, the external appearance of the blog and the neutrality of the information led to the human errors, and the blogger's information and the existence of the postal address of the products or services are the factors in the correct answers. Given that relative less research has investigated false positives or false negatives of human perceptions or decisions on information credibility research, we will identify and study the factors that strongly influence people's choices of incorrect answers by tackling the related results in the future research.

7 CONCLUSION

Through this work, we strove to connect the computational and human-centered approaches in understanding the credibility of the weblog post that is one of the primary sources of information acquisition and sharing. We found that, for the credibility evaluation on blog posts, humans consider four coherency factors (i.e., structure, alignment, effort text/image ratios) of the weblog post. We further experimentally proved the strong positive influence of the elements of coherency on building blog credibility models with respect to the both performance and efficiency. We hope that our study methods and results give researchers, developers, and practitioners a guideline for understanding and measuring the information credibility of weblogs (or broadly other social networking sites).

8 ACKNOWLEDGEMENTS

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (NRF-2017M3C4A7083529, NRF-2017R1C1B5017391).

REFERENCES

[1] Michael A Banks. 2008. *Blogging heroes: interviews with 30 of the world's top bloggers.* John Wiley & Sons.

[2] Mikhail Bautin, Lohit Vijayarenu, and Steven Skiena. 2008. International Sentiment Analysis for News and Blogs.. In *ICWSM*.

[3] Fabricio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virgilio Almeida. 2010. Detecting spammers on twitter. In *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)*, Vol. 6. 12.

[4] David Boud, Rosemary Keogh, and David Walker. 2013. *Reflection: Turning experience into learning.* Routledge.

[5] Evelyn M Boyd and Ann W Fales. 1983. Reflective learning: Key to learning from experience. *Journal of humanistic psychology* 23, 2 (1983), 99–117.

[6] Judee K Burgoon and Jerold L Hale. 1984. The fundamental topoi of relational communication. *Communication Monographs* 51, 3 (1984), 193–214.

[7] Brady Butterfield and Janet Metcalfe. 2006. The correction of errors committed with high confidence. *Metacognition and Learning* 1, 1 (2006), 69–84.

[8] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web.* ACM, 675–684.

[9] Thomas Chesney and Daniel KS Su. 2010. The impact of anonymity on weblog credibility. *International journal of human-computer studies* 68, 10 (2010), 710–718.

[10] Dietrich Dörner and Harald Schaub. 1994. Errors in planning and decision-making and the nature of human information processing. *Applied psychology* 43, 4 (1994), 433–453.

[11] Jonathan St BT Evans. 2008. Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.* 59 (2008), 255–278.

[12] Jonathan St BT Evans and Keith E Stanovich. 2013. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science* 8, 3 (2013), 223–241.

[13] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.

[14] Andrew J Flanagin and Miriam J Metzger. 2003. The perceived credibility of personal Web page information as influenced by the sex of the source. *Computers in human behavior* 19, 6 (2003), 683–701.

[15] BJ Fogg and Hsiang Tseng. 1999. The elements of computer credibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* ACM, 80–87.

[16] Brian J Fogg. 2003. Prominence-interpretation theory: Explaining how people assess credibility online. In *CHI'03 extended abstracts on human factors in computing systems.* ACM, 722–723.

[17] Brian J Fogg, Cathy Soohoo, David R Danielson, Leslie Marable, Julianne Stanford, and Ellen R Tauber. 2003. How do users evaluate the credibility of Web sites?: a study with over 2,500 participants. In *Proceedings of the 2003 conference on Designing for user experiences.* ACM, 1–15.

[18] Namrata Godbole, Manja Srinivasaiah, and Steven Skiena. 2007. Large-Scale Sentiment Analysis for News and Blogs. *Icwsm* 7, 21 (2007), 219–222.

[19] Kyungsik Han. 2018. How do you perceive this author? Understanding and modeling authors' communication quality in social media. *PloS one* 13, 2 (2018), e0192061.

[20] Thomas J Johnson and Barbara K Kaye. 2009. In blog we trust? Deciphering credibility of components of the internet among politically interested internet users. *Computers in Human Behavior* 25, 1 (2009), 175–182.

[21] Andreas Juffinger, Michael Granitzer, and Elisabeth Lex. 2009. Blog credibility ranking by exploiting verified content. In *Proceedings of the 3rd workshop on Information credibility on the web.* ACM, 51–58.

[22] Daniel Kahneman and Shane Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment* 49 (2002), 81.

[23] Andreas M Kaplan and Michael Haenlein. 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons* 53, 1 (2010), 59–68.

[24] Barbara K Kaye. 2007. Blog use motivations: An exploratory study. *Blogging, citizenship, and the future of media* (2007), 127–148.

[25] Gary A Klein. 2017. *Sources of power: How people make decisions.* MIT press.

[26] KP Krishna Kumar and G Geethakumari. 2014. Detecting misinformation in online social networks using cognitive psychology. *Human-centric Computing and Information Sciences* 4, 1 (2014), 14.

[27] Kyumin Lee, James Caverlee, and Steve Webb. 2010. Uncovering social spammers: social honeypots+ machine learning. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval.* ACM, 435–442.

[28] Stephan Lewandowsky, Ullrich KH Ecker, Colleen M Seifert, Norbert Schwarz, and John Cook. 2012. Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest* 13, 3 (2012), 106–131.

[29] Fangtao Li, Minlie Huang, Yi Yang, and Xiaoyan Zhu. 2011. Learning to identify review spam. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, Vol. 22. 2488.

[30] Yuqing Lu, Lei Zhang, Yudong Xiao, and Yangguang Li. 2013. Simultaneously detecting fake reviews and review spammers using factor graph model. In *Proceedings of the 5th annual ACM web science conference.* ACM, 225–233.

[31] Universal McCann. 2008. Universal McCann social media tracker wave 3, March 2008. *Universal McCann, New York* (2008).

[32] Meredith Ringel Morris, Scott Counts, Asta Roseway, Aaron Hoff, and Julia Schwarz. 2012. Tweeting is believing?: understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 conference on computer supported cooperative work.* ACM, 441–450.

[33] Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie S Glance. 2013. What yelp fake review filter might be doing?. In *ICWSM.* 409–418.

[34] Don Norman. 2013. *The design of everyday things: Revised and expanded edition.* Constellation.

[35] Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1.* Association for Computational Linguistics, 309–319.

[36] Rebecca B Rubin and Michael P McHugh. 1987. Development of parasocial interaction relationships. (1987).

[37] Richard Samuels. 2006. The magical number two, plus or minus: Some comments on dual-process theories. In *Proceedings of the 28 th Annual Conference of the Cognitive Science Society*, Vol. 202. 207.

[38] Herbert A Simon. 1956. Rational choice and the structure of the environment. *Psychological review* 63, 2 (1956), 129.

[39] Hyeonjin Soh. 2013. Attributes of trusted blog contents: through analysis of product-reviews in powerblogs and consumer survey. *The Journal of the Korea Contents Association* 13, 1 (2013), 73–82.

[40] Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. 2010. Detecting spammers on social networks. In *Proceedings of the 26th annual computer security applications conference.* ACM, 1–9.

[41] Peter M Visscher, Jian Yang, and Michael E Goddard. 2010. A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al.(2010). *Twin Research and Human Genetics* 13, 6 (2010), 517–524.

[42] Chih-Chien Wang and Hung-Yu Chien. 2012. Believe or Skepticism? An Empirical Study onIndividuals' Attitude to Blog Product Review. *International Journal of Innovation, Management and Technology* 3, 4 (2012), 343.

[43] Yingxu Wang and Vincent Chiew. 2010. On the cognitive process of human problem solving. *Cognitive systems research* 11, 1 (2010), 81–92.

[44] Debbie Weil. 2006. *The corporate blogging book*. Piatkus.

[45] Sung-Un Yang and Joon Soo Lim. 2009. The effects of blog-mediated public relations (BMPR) on relational trust. *Journal of Public Relations Research* 21, 3 (2009), 341–359.

[46] Yiming Yang and Jan O Pedersen. 1997. A comparative study on feature selection in text categorization. In *Icml*, Vol. 97. 412–420.