Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 55 (2019) 77–81

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Human-machine interaction: A case study on fake news detection using a backtracking based on a cognitive system ☆

Hoon Ko [a,1], Jong Youl Hong [b], Sangheon Kim [c], Libor Mesicek [d], In Seop Na [e,*]

[a] IT Research Institute, Chosun University, Gwangju, South Korea
[b] Hankuk University of Foreign Studies, Seoul, South Korea
[c] Department of History and Historical Content, Sangmyung University, Seoul, South Korea
[d] Jan Evangelista Purkyne University, Moskevska 54, 400 96 Usti nad Labem, Czech Republic
[e] Software Convergence Education Institute, Chosun University, Gwangju, South Korea

Received 16 November 2018; received in revised form 14 December 2018; accepted 27 December 2018
Available online 8 January 2019

Abstract

Although the Internet provides a variety of news, it also can give confusion caused by personal subjective thoughts such as personal TV, blogs, and unproven news. The unproven news is written in a subjective direction with added personal opinions rather than objective content, so readers may acquire knowledge with the wrong outlook. In addition, fake news is being produced and the problem of social polarization is becoming serious. In the end, it is necessary to detect the fake news, but it is not easy to distinguish the truth of published news because of the lack of fake news distinction time compared to the speed of information sharing on the Internet and the diversity and strong subjectivity of news. Therefore, in this paper, the possibility of fake news is defined by using the reverse-tracking method of the articles which are posted on the Cognitive System. Finally, as the result, the detection rate is average 85%.

2010 MSC: 00-01; 99-00

Keywords: Reverse-tracking; Cognitive system; Fake news; PageRank

1. Introduction

In 2016, the Oxford Dictionary selected 'post-truth' as the word of the year. In recent years, the issue of fake News has attracted a great deal of attention in the world through the US presidential election and Brexit, which is the political goal of individuals, parties, and organizations in the sense that the UK withdraws from the European Union. The New York Times reported that Cameron Harris made and distributed the fake news and earned $5000 for a short time, shared by six million people. Because of the nature of the Internet, the number of visits to celebrity sites is related to advertising revenue, so publishers constantly raise exciting false facts to increase the number of visitors (Contributors, 2018a; Ogiela & Ogiela, 2014). As fake news is read by more and more people, it is affecting many people who are interested in politics. Harris' fake news can be spread out through social networks, or even searched through the Google search engine. More fake news is produced because visitors accessing the fake news website

☆ Fully documented templates are available in the elsarticle package on http://www.ctan.org/tex-archive/macros/latex/contrib/elsarticleCTAN.

* Corresponding author.
E-mail addresses: skoh21@chosun.ac.kr (H. Ko), herr_hong@hanmail.net (J.Y. Hong), libor.mesicek@seznam.cz (L. Mesicek), ypencil@hanmail.net (I.S. Na).
[1] First Author.

https://doi.org/10.1016/j.cogsys.2018.12.018
1389-0417/

become an ad base. On the other hand, the fake news in Korea is more of a political dispute, rather than an income source from advertising. In other words, making fake news about a politician that they do not support helps reinforce their solidarity. In the case of Korea, during the political turmoil in the second half of 2016, there was social controversy due to the fake news, and broadcast stars suffered from the fake news (Sangaiah, Samuel, Li, Abdel-Basset, & Wang, 2018). There are penalties for the production and sharing of the fake news both at home and abroad, but it is impossible to completely prevent the production and spread of the fake news and it takes a lot of time to distinguish between them (Han, 2017). In the end, it is necessary to determine the truth of a posting considering the time of posting, the author's relationship with other posts, the credibility of the author, and where they share it, but it is impossible for a person to handle it all. Therefore, a system that defines the reliability of the news is needed and the system should be able to grasp the surrounding situation and the characteristics of the news. This study suggests a cognitive system applying backtracking to judge false news based on the characteristics of web posting (Contributors, 2018c). To date, the characteristics of the fake news are that people with malicious intent post the fake news and others following the malice share the news (Contributors, 2018b). Reverse-Tracking on the fake news can reveal features that focus on a certain few users, rather than passing similar news from a variety of sources (Ogiela, 2008; Pang & Goodwin, 2003). If it were a normal issue article, similar posts would have been found in various places. According to this research, the provider of the news posted in the place as the personal blog is backtracked. Through such a process, we identify the characteristics of the first published place (such as newspaper, bureau, freelance journalist, etc.). At the same time, we understand the diversity of sites that have published the topic. Through such a combination of methods, the possibility of false news is determined. In other words, we propose cognitive system based on backtracking. The composition of this study is as follows. Chapter 2 summarizes the related research that defines fake news and cognitive system, and chapter 3 explains the proposed Fake News Detect System (FNDS) based on Reverse-Tracking with a cognitive system. Chapter 4 analyzes and summarizes the use cases using the proposed method, and chapter 5 concludes.

2. The related work

2.1. Fake news case

Recently 'Empire News' and 'National Report' have been suspected of deliberately creating the fake news and taking economic advantage (Tognotti, 2016). In the 2016 US presidential election, more and more people were producing the fake news that their supporters wanted, in the process of driving the mainstream media into false and distorting journalism in camps (Contributors, 2018a). In the case of Korea in 2016, the fake news was raised as a new social issue in response to the impeachment of the president and early election campaign. Several media sourced pointed out the problem of the fake news. Through the fact check, they revealed the real news and the fake news. However, the opposing camp also used false evidence to claim that the news was the fake news. It was a socially confusing situation (Han, 2017; Ogiela & Ogiela, 2016).

2.1.1. Explanation

When an event occurs, it can be interpreted differently depending on the viewpoint of the person facing the event. The person can also post what he thinks on his blog. This is not the fake news, but a summary of his thoughts. The Fake news, however, creates artificial events for the purpose of harming certain people. Ordinary people read the text and believe it without doubt, and if they have the same or similar texts elsewhere, they take it seriously.

Assumption. Articles about issues are displayed on various sites at the same time, not at one site.

Explanation. There are many media companies around the world, including newspapers, broadcasters, and freelance journalists. If there are big social issues, they post news at the same time. Readers see published articles and share them on their blogs. At this time, the number of times of sharing is different according to the preference of the media company or the reporter's blog. However, the number of times a certain company or blog is shared does not increase unilaterally. Eventually, those who produce fake news continue to produce the fake news and those who follow them share the fake news unilaterally.

2.1.2. News sharing

There are page ranking methods that rank web pages based on web graphs as a way as to measure how many visitors there are (Cicone & Serra-Capizzano, 2010; Markus Sobek, 2006). Fig. 1 is a shared process of generated news. (a) is a process of sharing the news about actual events. If any event occurs, many media sources will simultaneously write articles about the event. Those who see the articles of each medium will share the articles of the medium they want. (b) shows the process of sharing about a bogus event. When one person creates the fake news and other people closely associated with that person share the news, they continue to share to new people continuously. Of course, depending on the nature of the web, regardless of (a) or (b), anyone who wants to share can share articles on other webs or other media that have already been shared (Red Color: Crossed Shared). Eq. (1) summarizes the formula for measuring the relative importance of pages by analyzing outlink information of Page Rank page (Markus Sobek, 2006).

$$PR(A) = (1-d) + d\left(\frac{PR(T1)}{C(T1)} + \frac{PR(T2)}{C(T2)} + \cdots + \frac{PR(Tn)}{C(Tn)}\right) \quad (1)$$

2.2. Difficulty in distinguishing fake news

Even if you are exposed to fake news using Facebook, Twitter, or blogs, it is hard to tell whether the information is fake or genuine. Even though a related organization professes the truth about the news, it is difficult to distinguish it, unless it is confirmed directly by going to the site. Most of the public does not know which site to identify as the fake news.

3. FNDS based reverse-tracking

3.1. Backtracking

When a third party shares the news shared on the web and reopens it, it is difficult to know whether the news is true or false. However, if you trace back from where you share it, you can analyze where you share the source news from (Li, Ding, Hu, & Zheng, 2018; Pang & Goodwin, 2003; SJ, 2015). Fig. 1(a), in the case of Normal News, $S_{11.x} \rightarrow S_{111} \rightarrow S_{11} \rightarrow S_1 \rightarrow G$ (n, 1) or $S_{12.y} \rightarrow S_{121} \rightarrow S_{12} \rightarrow S_2 \rightarrow G$ (n,1). With this analysis, we can see that news of G (n, 1) is shared by $S_{11\ldots x}$ and $S_{12\ldots x}$. Table 1 shows the results obtained by the same method.

Algorithm 1. FNDS Reverse-Tracking Algorithm

Require: $INPUT(M) \leftarrow [S_1, S_2, \ldots, S_x]$
Require: $Reverse_Tracking(S.root) \leftarrow S_{(1,2,\ldots,x)}$
for do$(i = 0 : i < MAX; i++)\{$
while $S_i = S_x$ **do**
$Add(Matched\_Site.List) \leftarrow S_i$
$matched\_count \leftarrow matched\_count + 1,$
end while
while $S_i \neq S_x$ **do**
$Add(nomatched\_Site.List) \leftarrow S_i$
$nomatched\_count \leftarrow nomatched\_count + 1\}$
end while
$\}$
$Result = \frac{matched\_count}{matched\_count + nomatched\_count}$
$PrintResult$
if $S_i == S_{1,\,2,\,\ldots,\,x}$ **then** $\{$
$DECISION(FAKE\_NEWS)$
OR
$DECISION(REAL\_NEWS)$
$\}$

The algorithm explains the FNDS Reverse-Tracking System that it suggested. The sensors detect the messages and forwards it to Learning & Reasoning to understand what it is. In Learning & Reasoning Step, to try to know if it is the fakenews, it has to send the message to FNDS which contains 'Dummy news detection' (Navarro et al., 2018) and 'Backtracking' (Anni & Sangaiah, 2018). In Dummy news detection, it doesn't know if it is real news, so just it considers them as a dummy news. If they are dummy news, it puts them into Reverse-Tracking step. In Reverse-Tracking step, it traces to where they are from. According to the result from Reverse-Tracking process, it decides Fake News or Real News, then it returns the result to Learning & Reasoning, next it shares the results with Knowledge & Models and Planning & Cognitive Control. Finally, it effects users' action with the results.

In other words, the same news is generated (posted) by $G_{n(1,\,\ldots?)}$. The intermediate sharers share the news from where their preferences are high, and the next sharer shares in the same way. After all, it's the same news, but the Reverse-Tracking where you share it points to different routes. On the other hand, the results of reverse tracking of the fake news are summarized in Table 1. As mentioned in 3.1, the first news generation is assembled into $G_{(n)}$.

In the case of Table 1, it is common for many media companies to often report and post at the same time when an event occurs. In Table 1, in the case of the fake news, it is interpreted as an artificially maneuvered posting, which is shared by other followers on an ongoing basis. Therefore, it can be defined that the probability of the fake news is high in Table 2.

Proof: If you trace back to where the sources of the news appear on blogs shared by followers and those who re-shared the shared news, point to one. The Google Page Rank algorithm is applied to how to determine fake news.

4. Experiment and analysis

Based on Chapter 3, we analyzed news posted on the web. In the analysis method, first, key words were defined and analyzed by 1–4 which are from fake news in [Table 3] and real news in [Table 4].

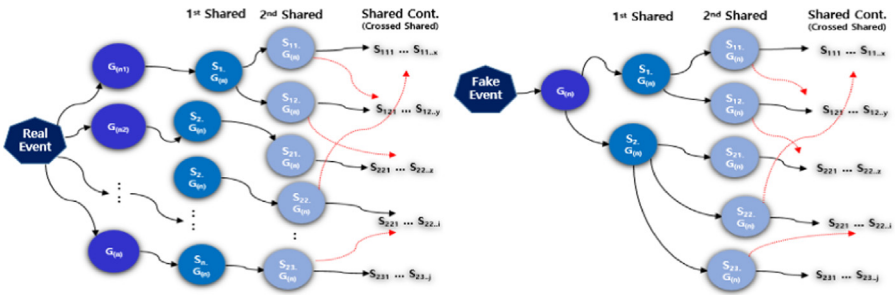
4.1. Case study 1: Fake news

News 1, posted in February 9th, 2017. An article about the blacklisting of extreme right movie studios and Kim Jong Dae, a member of the National Assembly.

Order of priority 1. Justice Party, Kim Jong Dae, film studio, blacklist.

⟨ Words Sequence ⟩

- When one or two words were used, the efficiency of the analysis was not good. Three words and four words were combined and analyzed.
- We prioritized the important keywords of the news, and analyzed three words and four words for web search.
- Priority words must be included in three or four words.
- 3 Words: Justice Party → Kim Jong Dae → Film Studio.
- 4 Words: Justice Party → Kim Jong Dae → Film Studio → Blacklist.



(a) Normal Event (b) Fake Event

Fig. 1. Normal event vs fake event.

Table 1
Reverse-tracking of real news.

News generation (G_n)	Final sharer	Intermediate sharer
$G_{(n,1)}$	$S_{11...x}$	$S_1.G_{(n)} \rightarrow S_{11}.G_{(n)} \rightarrow S_{111}$
	$S_{12...x}$	$S_2.G_{(n)} \rightarrow S_{12}.G_{(n)} \rightarrow S_{121}$
$G_{(n,2)}$	$S_{22...z}$	$S_2.G_{(n)} \rightarrow S_{21}.G_{(n)} \rightarrow S_{221}$
$G_{(n,?)}$	$S_{22...j}$	$S_2.G_{(n)} \rightarrow S_{22}.G_{(n)} \rightarrow S_{221}$
	$S_{23...j}$	$S_n.G_{(n)} \rightarrow S_{23}.G_{(n)} \rightarrow S_{231}$

Table 2
Reverse-tracking of fake news.

News generation (G_n)	Final sharer	Intermediate sharer
$G_{(n,?)}$	$S_{11...x}$	$S_1.G_{(n)} \rightarrow S_{11}.G_{(n)} \rightarrow S_{111}$
	$S_{12...x}$	$S_2.G_{(n)} \rightarrow S_{12}.G_{(n)} \rightarrow S_{121}$
	$S_{22...z}$	$S_2.G_{(n)} \rightarrow S_{21}.G_{(n)} \rightarrow S_{221}$
	$S_{22...i}$	$S_2.G_{(n)} \rightarrow S_{21}.G_{(n)} \rightarrow S_{221}$
	$S_{22...j}$	$S_2.G_{(n)} \rightarrow S_{21}.G_{(n)} \rightarrow S_{221}$

4.2. Case study 2. Real news

News 2, posted in September xx, 2017. An article about North Korea, nuclear test, earthquake, Kim Jeong Eun. In 2017, North Korea conducted an underground nuclear experiment and soon afterwards an earthquake occurred.

Order of priority 1. North Korea, Nuclear Experiment, Earthquake, Kim Jeong Eun.

⟨ Words Sequence ⟩

- 3 Words: North Korea → Nuclear Experiment → Earthquake.
- 4 Words: North Korea → Nuclear Experiment → Earthquake → Kim Jeong Eun.

5. Conclusion

Fake news is reported to the related organization when the reader reads the news and becomes suspicious. Next, the organization decides the authenticity of the news and shares the fake news information. While this process has the advantage of being determined carefully, it takes a lot of time to decide on it as being fake news. The proposed method can be judged faster than the existing method. For example, if there is a real social incident, all the media will cover the case and post the article on the web. The arti-

Table 3
Case study 1. Fake news with 3 words and 4 words.

Words bundle	Analysis	Number of news source
3 Words: Justice party → Kim Jong Dae → Film studio	(1) Search results found in a large number of sites: 20 or more, (2) 9 of 10 sites shared Nocutilbe	(1) Number of shared roots: 1, (2) Detect Rate: 90%
4 Words: Justice Party → Kim Jong Dae → Film studio → Blacklist	(1) 9 retrieved from 3 words, (2) Two less are searched, (3) Search count decreased from 10 to 8, (4) 7 out of 8 included 4 words, (5) 7 of 8 shared Nocutilbe, (6) Search rate-Down, the number of search-Down	(1) Number of shared roots: 1, (2) Detect Rate: 87.5%

Table 4
Case study 2. Real news with 3 words and 4 words.

Words bundle	Analysis	Number of news source
3 Words: North Korea → Nuclear test → Earthquake	(1) 8 out of 10 articles are from different sites. (2) 2 articles almost brought the same article. So it's hard to see them as a unique site/source	Detect Rate: 80%
4 Words: North Korea → Nuclear test → Earthquake → Kim Jeong Eun	The list of 10 shows that they all use similar articles on various media	Detect Rate: 80%

cles of each media company will be different, but the keywords will be included at the same time, and the secondary media or social network users will share articles of the media that they frequently visit. At this time, if you trace social users' articles in reverse, you can see that they are shared by various media companies. However, the person generating the fake news generates the false news for a specific purpose, and the social users who follow it share the news. Just as with real news analytic, reverse tracking of shared articles leads to one or a few websites, unlike real news. Based on these algorithms, we analyzed the news in 2017. As a result of the analysis, the article decided to be fake news had one news producer. Although it suggested new methods the fake news, still there are some limitations. There are many journalists who belong to a company or free-lancer and they posts some news with their idea. Surely the posts are based on the real-accidents, but the posts will be based on the journalists' philosophy, then it can be a subjective post. With the reverse-tracking system, it is hard to detect between the fake news and the subjective posts. In the future, how to detect between the fake news and the subjective posts should be included in the advanced system.

Acknowledgements

This study was supported by research fund from Chosun University, 2018.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cogsys.2018.12.018.

References

Anni, J. S., & Sangaiah, A. K. (2018). Wireless integrated sensor network: Boundary intellect system for elephant detection via cognitive theory and fuzzy cognitive maps. *Future Generation Computer Systems, 83*, 522–534.

Cicone, A., & Serra-Capizzano, S. (2010). Google pageranking problem: The model and the analysis. *Journal of Computational and Applied Mathematics, 234*(11), 3140–3169.

Contributors, W. (2018a). Alternative facts — Wikipedia, the free encyclopedia [Online]. Retrieved from https://en.wikipedia.org/w/index.php?title=Alternative_facts&oldid=866156336. Accessed 13-November-2018.

Contributors, W. (2018b). Fake news — Wikipedia, the free encyclopedia. Retrieved from https://en.wikipedia.org/w/index.php?title=Fake_news&oldid=867877470.

Contributors, W. (2018c). Nineteen eighty-four - Wikipedia, the free encyclopedia [Online]. Retrieved from https://en.wikipedia.org/w/index.php?title=Nineteen_Eighty-Four&oldid=868532366. Accessed 13-November-2018.

Han, S. (2017). Fake news — slownews, fake news (korean version) [Online]. Retrieved from http://slownews.kr/62053. Accessed 27-February-2017.

Li, B., Ding, L., Hu, D., & Zheng, S. (2018). Back tracking algorithm-based disassembly sequence planning. *Procedia CIRP, 69*, 932–937.

Markus Sobek (2006). The pageranking algorithm — eFactory GmbH & Co.KG. Retrieved from https://pr.efactory.de/e-pagerank-algorithm.shtml.

Navarro, J., Doctor, F., Zamudio, V., Iqbal, R., Sangaiah, A. K., & Lino, C. (2018). Fuzzy adaptive cognitive stimulation therapy generation for alzheimer's sufferers: Towards a pervasive dementia care monitoring platform. *Future Generation Computer Systems*.

Ogiela, L. (2008). Cognitive computational intelligence in medical pattern semantic understanding. *Fourth international conference on natural computation, 2008. icnc'08* (Vol. 6, pp. 245–247). IEEE.

Ogiela, L., & Ogiela, M. R. (2014). Data mining and semantic inference in cognitive systems. In *2014 international conference on intelligent networking and collaborative systems (incos)* (pp. 257–261). IEEE.

Ogiela, M. R., & Ogiela, L. (2016). On using cognitive models in cryptography. In *Advanced information networking and applications (AINA). 2016 IEEE 30th international conference on* (pp. 1055–1058). IEEE.

Pang, W., & Goodwin, S. D. (2003). A graph based backtracking algorithm for solving general csps. In *Conference of the canadian society for computational studies of intelligence* (pp. 114–128). Springer.

Sangaiah, A. K., Samuel, O. W., Li, X., Abdel-Basset, M., & Wang, H. (2018). Towards an efficient risk assessment in software projects-fuzzy reinforcement paradigm. *Computers & Electrical Engineering, 71*, 833–846.

SJ (2015). Introduction to backtracking programming —Algorithms [Online]. Retrieved from https://algorithms.tutorialhorizon.com/introduction-to-backtracking-programming/. Accessed 16-March-2015.

Tognotti, C. (2016). 9 fake news sites that are spreading dangerous, yet believable stories. Retrieved from https://www.bustle.com/articles/195286-9-fake-news-sites-that-are-spreading-dangerous-yet-believable-stories.