Detecting Rumors Transformed from Hong Kong Copypasta

Yin-Chun Fung[1], Lap-Kei Lee[1], Kwok Tai Chui[1], Ian Cheuk-Yin Lee[1], Morris Tsz-On Chan[1], Jake Ka-Lok Cheung[1], Marco Kwan-Long Lam[1], Nga-In Wu[2] and Markus Lu[3]

[1] School of Science and Technology, Hong Kong Metropolitan University, Ho Man Tin, Kowloon, Hong Kong SAR, China
[2] College of Professional and Continuing Education, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China
[3] Hong Kong International School, Hong Kong SAR, China
ycfung@study.hkmu.edu.hk,{lklee,jktchui}@hkmu.edu.hk, ngain.wu@cpce-polyu.edu.hk

Abstract. A copypasta is a piece of text that is copied and pasted in online forums and social networking sites (SNSs) repeatedly, usually for a humorous or mocking purpose. In recent years, copypasta is also used to spread rumors and false information, which damages not only the reputation of individuals or organizations but also misleads many netizens. This paper presents a tool for Hong Kong netizens to detect text messages that are copypasta or their variants (by transforming an existing copypasta with new subjects and events). We exploit the Encyclopedia of Virtual Communities in Hong Kong (EVCHK), which contains a database of 315 commonly occurred copypasta in Hong Kong, and a CNN model to determine whether a text message is a copypasta or its variant with an accuracy rate of around 98%. We also showed a prototype of a Google Chrome browser extension that provides a user-friendly interface for netizens to identify copypasta and their variants on a selected text message directly (e.g., in an online forum or SNS). This tool can show the source of the corresponding copypasta and highlight their differences (if it is a variant). From a survey, users agreed that our tool can effectively help them to identify copypasta and hence help stop the spreading of this kind of online rumor.

Keywords: rumor detection, copypasta, natural language processing

1 Introduction

Nowadays, many express their opinions by publishing or reposting articles on the Internet. According to Leung [1], online discussion forums are the preferred social medium for gaining recognition; blogs and social networking sites (SNSs) like Facebook are normally used for social and psychological needs and the need for affection. The popularity of these online platforms has also led to the rapid spreading of rumors, which is false information created by, for example, exaggeration,

tampering, or mismatching, and can mislead the readers and even have a negative impact on public events.

A copypasta is a piece of text that is copied and pasted repeatedly around the Internet and can usually be seen in online discussion forums and social networking sites for humorous or mocking purposes [3, 4]. Some articles may be transformed or adapted from existing copypasta by replacing the subjects and/or events with new ones. These copypastas (and their variants) are usually funny or satirical, and this can bring happiness to their readers. Yet these copypastas may evolve into rumors because some true believers would believe that the content of the copypasta is true [5]. For example, in November 2019, a copypasta "A fierce fight broke out at the top of the government", which is transformed from the copypasta about a fight between Netherlands national football team players in 2012, appeared in a popular forum "LIHKG" in Hong Kong, and the then Chief Secretary for Administration of the Hong Kong government dismissed the rumor on his official Facebook page [6]. Such a rumor can be regarded as rumors created by tampering [2]. Twitter, one of the popular SNSs, has also updated its security policies to combat false information caused by copypasta [7].

Copypasta may be easily identified by netizens who are frequent users of online forums and SNSs. Yet, many other netizens fail to identify copypastas from credible and authentic texts. These rumors are adapted from different articles and events or are made from imagination. This culture has become very common for netizens worldwide. Some netizens may think that the transformation to copypastas will make them more interesting. Some others however maliciously adapted them to achieve purposes like defamation. At present, there is no relevant law in Hong Kong to regulate this kind of behavior[1], so it is becoming more common in Hong Kong.

The Encyclopedia of Virtual Communities in Hong Kong (EVCHK)[2] is a website with a collection of more than 12,000 entries on the Internet culture in contemporary Hong Kong, and it is operated by a community of volunteer editors in a similar fashion to Wikipedia [4]. EVCHK contains a database of 315 common copypastas in Hong Kong, which may be helpful for netizens to find out manually whether a text is a copypasta, a transformed variant, or neither of them. Such checking of online articles or text messages however would require a lot of time and energy.

To mitigate the problem, one direction is to educate the netizens to raise their cybersecurity awareness, e.g., [8]; another direction is to develop tools that can automatically detect whether a text is a rumor. There have been many rumor detection algorithms in the Natural Language Processing (NLP) research community (see the surveys [9, 10] and the references therein). While sentiment analysis, e.g., [11, 12], and intent identification, e.g., [13, 14], are well-studied NLP problems and have many applications, e.g., chatbots [8, 15], rumor detection may involve the use of these and more textual features for the machine learning algorithms and most of the existing works can only identify a rumor without offering an explanation why it is a rumor.

---

1 https://www.info.gov.hk/gia/general/202003/18/P2020031800422.htm

2 https://evchk.fandom.com/

This paper aims to develop a tool that can identify copypasta (and its variant) in Hong Kong and provide explanations for why texts are identified as copypasta. Our contributions include the following:

- There is no publicly available dataset for copypasta detection. We collected copypasta (including their variants) and non-copypasta (i.e., text messages that are not copypasta) from different websites, including EVCHK, online forums, and news media in Hong Kong, and created a dataset for Hong Kong copypasta detection.
- We divided our dataset for training and testing, respectively. We trained machine learning models based on CNN and RNN, and found that CNN performs better on copypasta detection. The accuracy is around 98% on the testing dataset.
- We developed a prototype of a Google Chrome browser extension that provides a user-friendly interface for netizens to identify copypasta and their variants on a selected text message directly (e.g., in an online forum or SNS). This tool can show the source of the corresponding copypasta and highlight their differences (if it is a variant).
- We conducted a survey on 45 users and focus group interviews with some participants to show that our tool can effectively help them to identify copypasta and help stop the spreading of this kind of online rumor.

Organization of the paper. Section 2 reviews some existing works on rumor detection. Section 3 gives the detailed design of our copypasta detection tool. Section 4 presents an evaluation of the tool on 45 participants. Section 5 concludes the paper and proposes some future work directions.

2 Existing works

The Encyclopedia of Virtual Communities in Hong Kong (EVCHK) is one of the most popular websites for netizens to find out the meaning and reference of net slang and copypasta in Hong Kong, as information in EVCHK is well-organized into different categories by a community of volunteer editors. Users can often find the relevant pages in EVCHK using search engines like Google Search. Yet those with lower ICT literacy may not able to find the desired information using appropriate keywords using the search function in EVCHK and search engines.

Slang detection and identification. To the best of our knowledge, there is no existing work focusing on copypasta detection. A closely related work is the problem of slang detection and identification, proposed by Pei et al. [16]. They used RNN to identify the exact positions of slang keywords to detect the presence of slang in a sentence. However, it locates the slang in a sentence only and cannot be adapted to identify copypasta.

"Snopes" website[3]. It is a famous and popular website for users to fact-check the source of articles, news, and copypasta (see Fig. 1 (left)). It provides a Fact Check Rating System that is credible to users. However, it focuses on English content for US news media and websites, so users using other languages cannot utilize their fact-checking service.

HKBU FactCheck Service[4]. This service (see Fig. 1 (right)) is developed by the School of Communication of the Hong Kong Baptist University (HKBU). Like the "Snopes" website, it provides a Fact Check Rating System to fact-check the article and news and it includes more Hong Kong local articles and news. Users can make a request to fact-check an article on the website, but fact-checking is only done manually and would take a long time. It only has around 10 fact-checking results per month. Identifying copypasta is also not a focus of this service.

TweetCred. It is a tool developed by Gupta et al. [17], which is a Google Chrome browser extension that provides a credibility rating for each tweet. The tool uses a supervised automated ranking algorithm to evaluate the credibility of a tweet such that users can determine whether a tweet is a rumor or not. However, it can only be used for tweets on Twitter and cannot identify copypasta.

Fig. 1. Screen Captures of the Snopes website (left) and HKBU FactCheck Service (right).

3 https://www.snopes.com/
4 https://factcheck.hkbu.edu.hk/

Table 1. Comparison of existing tools on Hong Kong copypasta detection.

Solution	Snopes websites	Slang detection [16]	HKBU FactCheck Service	TweetCred	Our tool
Detects Hong Kong copypasta	✗	✗	✔	✗	✔
Provides automatic detection	✗	✔	✗	✔	✔
Shows the source of the copypasta	✗	✗	✗	✗	✔

3 Design of Our Copypasta Detection Tool

3.1 Overall system architecture

Our solution can be split into two parts: the server side and the client side. The client side is a JavaScript extension that makes API calls to the server side, which is a web service written in Python.

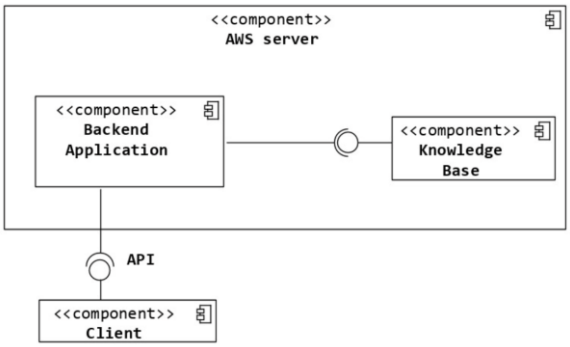

Fig. 2. Overall system architecture of our tool.

Figure 2 shows the overall system architecture of our tool. The Google Chrome browser extension (i.e., the Client) gets the text selected by the user and sends an API request to the server. The server contains a web service (the backend application) which has an underlying machine learning detection model trained with a knowledge base of copypasta and non-copypasta to predict the probability that a text is a copypasta (or its variant) and then provide the source of the copypasta if the text is identified as a copypasta. The browser extension receives the result from the web service and then displays the analysis result to the user.

Figure 3 shows the workflow of the browser extension. The text input can be text passages selected by the user on the browser (which is referred to as *user post*) or text

inputted by the user directly on the browser extension (which is referred to as *user input*). We employed a very simple text preprocessing strategy by removing all non-Chinese characters and symbols (e.g., punctuation marks and emojis) on the user post and user input.

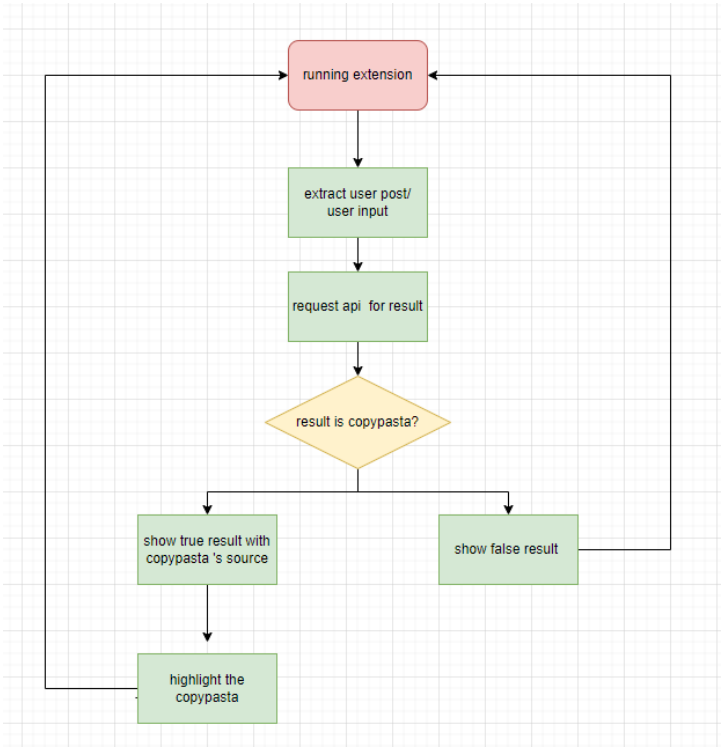

Fig. 3. Workflow of the browser extension.

3.2 The machine learning models

The detection model is trained using neural networks. Two models have been tested and compared (the comparison result is given in Section 4).

The first model is a convolutional neural network (CNN), which consists of the word embedding, convolution, pooling, and output layers. The word embedding layer is used to make the dimension vector of the data dictionary. The convolution layer is for detecting features. The pooling layer is for reducing data dimensions and for generalizing the feature. The output is the fully connected layer with dropout and SoftMax output.

The second model is a recurrent neural network (RNN). Long Short-Term Memory (LSTM, which is a type of RNN) was adopted. The neural network consists of an embedding, a fully connected layer, a sequence pool, and an output layer. There are one LSTM operation and two max sequence pool operation. The output layer is using SoftMax with a size of 2.

3.3 User interface of the browser extension

As the tool is designed for Hong Kong netizens and most of them use Cantonese (a dialect of the Chinese language), the browser extension uses only Cantonese for the user interface. Once the browser extension is installed on the Google Chrome browser, we can select some text passages and then right-click on the selected text to detect whether the selected text is copypasta or not (Fig. 4). Alternatively, we can start the browser extension directly on the browser's menu bar, which displays a text box for user text input (Fig. 5 (left)) and a submit button. The detection result page is shown in Fig. 5 (right). A detection summary with the copypasta probability is shown. There are three levels of probability: definitely a copypasta (≥ 80%), maybe a copypasta (≥ 50% and < 80%), and not a copypasta (< 50%). For the first two levels, a link to the EVCHK page for the source of the copypasta will be given right below the probability, and the text that matches the copypasta source will be highlighted at the bottom of the result page.

Fig. 4. Right-click on the selected text to detect copypasta using the browser extension.

Fig. 5. User interface of the browser extension: User input (left) and detection result (right).

4 Evaluation

Our evaluation focuses on the performance of the machine learning models and evaluation on 45 users.

4.1 Performance of the machine learning models

Our created dataset. We gathered 233 out of the 315 copypastas from EVCHK as the source. We have collected 15,963 texts from Google News and the popular Hong Kong online forum LIHKG[5] using web scraping techniques and classified them into copypasta and non-copypasta with references to the sources.

Table 2. No. of text message samples in our created dataset.

Item	Training	Testing	Total
Copypasta	409	288	697
Non-copypasta	5,446	10,020	15,266
Source of copypasta	-	-	233

Comparisons between CNN and RNN. We trained a CNN model and an RNN model using the same training dataset and performed comparison and evaluation of their performance using the same testing dataset. Table 3 shows the results. Both models have high accuracy and recall, but the CNN model is better than the RNN model on all indicators. As most of the text is non-copypasta in the dataset, precision is more concerned. Since the precision of the CNN model is 0.6911, it is regarded as having a better performance in detecting copypasta.

Table 3. Training results with different models.

	CNN model	RNN model
True Positive	273	268
True Negative	9898	9772
False Positive	122	248
False Negative	15	20
Accuracy	0.9867	0.9740
Precision	0.6911	0.5194
Recall	0.9479	0.9306
F1-score	0.7994	0.6667

Comparisons on text preprocessing strategies. We also conducted an experiment on text preprocessing strategy will yield a better result. The strategies are: (1) no preprocessing, (2) remove stop words, and (3) keep only Chinese characters. We used PyCantonese [18] to remove the stop words in Strategy (2). As shown in Table 4, Strategy (3) gives a higher F1 score.

[5] https://lihkg.com/

Table 4. Training results with different text preprocessing strategies

	no preprocessing	**remove stop words**	**keep only Chinese characters**
True Positive	275	272	273
True Negative	9871	9850	9898
False Positive	149	170	122
False Negative	13	16	15
Accuracy	0.9843	0.9820	0.9867
Precision	0.6486	0.6154	0.6911
Recall	0.9549	0.9444	0.9479
F1-score	0.7725	0.7452	0.7994

4.2 User evaluation

We invited 45 undergraduate students in Hong Kong to test our tool and complete a Google Form survey. All participants are experienced netizens and native Cantonese speakers. The survey consists of 6 questions on a 5-point Likert scale (1: disagree, 2: partially disagree, 3: neutral, 4: partially agree, 5: agree). Table 5 shows the survey results.

Table 5. Survey result

Item	1	2	3	4	5
1. Our user interface is attractive.	0 (0%)	6 (13.3%)	3 (6.7%)	11 (24.4%)	25 (55.6%)
2. The extension is useful for determining whether a text passage is a copypasta.	1 (2.2%)	2 (4.4%)	12 (26.7%)	14 (31.1%)	16% (35.6%)
3. The extension can help me identify rumors or fake news.	0 (0%)	5 (11.1%)	15 (33.3%)	14 (31.1%)	11 (24.4%)
4. The extension detection result returns efficiently related information when the text is a copypasta.	0 (0%)	5 (11.1%)	9 (20%)	15 (33.3%)	16 (35.6%)
5. The extension increases my awareness of copypasta.	0 (0%)	11 (24.4%)	7 (15.6%)	12 (26.7%)	15 (33.3%)
6. I would recommend this extension to others.	0 (0%)	6 (13.3%)	12 (26.7%)	13 (28.9%)	14 (31.1%)

In the survey, more than half of the participants gave positive feedback on the user interface design (Question 1). Our tool receives an average mark of 4 in Questions 2 to 4, which reflects that the tool is useful for the participants to identify copypasta and rumors. The majority of participants agree that the tool increases their awareness of copypasta and would recommend it to others.

Among the 45 participants, we invited 4 of them to join a focus group interview so as to get more qualitative responses from them. They think the extension is creative because they have never seen an extension that detects the copypasta before. However, they also found that the tool cannot detect non-Chinese copypasta and Chinese copypastas translated from other languages which can be a future work direction. They also expressed concerns that the tool cannot detect the new copypasta, as it requires human effort to update the knowledge base on the server side.

5 Conclusion and Future Work

In this paper, we collected copypasta and non-copypasta from different websites, forums, and news in Hong Kong and created a dataset for Hong Kong copypasta. We trained CNN and RNN prediction models and found that CNN performs better than RNN in copypasta detection. We also showed how the tool can be implemented to facilitate usage: The prediction model is deployed to a server, and a Google Chrome extension can be designed to communicate with the server for users to detect whether selected text is a copypasta or not, and get more information about the identified copypasta (if any). A user survey of our tool showed that our detection tool is straightforward to use and is useful to detect copypasta. Users can learn more about Hong Kong copypasta and reduce their chance of believing some rumors from copypasta. This would help stop the spreading of online rumors from copypasta.

Limitations and future works. Our tool cannot detect new copypasta that is not in our dataset. As a future work direction, the extension may provide a reporting feature for users to report new copypastas when they found them missing in our detection result. When the dataset is sufficient, the model can be trained to classify the source. Another future work direction is to make the extension fully automatic; the extension may work with a script running in the background that grabs the text of the web page and checks whether they are copypasta automatically.

References

1. Leung, L.: Generational differences in content generation in social media: The roles of the gratifications sought and of narcissism. Computers in Human Behavior 29(3), 997-1006 (2013).
2. Chen, J., Wu, Z., Yang, Z., Xie, H., Wang, F. L., Liu, W.: Multimodal fusion network with contrary latent topic memory for rumor detection. IEEE MultiMedia 29(1), 104-113 (2022).
3. Riddick, S., Shivener, R.: Affective spamming on Twitch: Rhetorics of an emote-only audience in a presidential inauguration livestream. Computers and Composition 64, 102711 (2022).
4. Lam, C.: How digital platforms facilitate parody: Online humour in the construction of Hong Kong identity. Comedy Studies 13(1), 101-114 (2022).
5. Zannettou, S., Sirivianos, M., Blackburn, J., Kourtellis, N.: The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. Journal of Data and Information Quality 11(3), 1-37 (2019).

6. Facebook page of the Hong Kong Chief Secretary for Administration's Office, https://www.facebook.com/CSOGOV/posts/420752431929437, last accessed 2022/07/01.
7. Avery, D. Twitter updates security policy to combat spam tweets and 'copypasta', https://www.cnet.com/news/social-media/twitter-updates-security-policy-to-combat-spam-tweets-and-copypasta/, last accessed 2022/07/01.
8. Fung, Y. C., Lee, L. K.: A chatbot for promoting cybersecurity awareness. In: Agrawal, D. P., Nedjah, N., Gupta, B. B., Martinez Perez, G. (eds.) Cyber Security, Privacy and Networking, LNNS, vol. 370, pp. 379-387. Springer, Singapore (2022).
9. Meel, P., Vishwakarma, D. K.: Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Systems with Applications 153, 112986 (2020).
10. Rani, N., Das, P., Bhardwaj, A. K.: Rumor, misinformation among web: A contemporary review of rumor detection techniques during different web waves. Concurrency and Computation: Practice and Experience 34(1), e6479 (2022).
11. Fung, Y. C., Lee, L. K., Chui, K. T., Cheung, G. H. K., Tang, C. H., Wong, S. M. Sentiment analysis and summarization of Facebook posts on news media. In: Data Mining Approaches for Big Data and Sentiment Analysis in Social Media, pp. 142-154. IGI Global (2022).
12. Lee, L. K., Chui, K. T., Wang, J., Fung, Y. C., Tan, Z.: An improved cross-domain sentiment analysis based on a semi-supervised convolutional neural network. In: Data Mining Approaches for Big Data and Sentiment Analysis in Social Media, pp. 155-170. IGI Global (2022).
13. Liu, Y., Liu, H., Wong, L. P., Lee, L. K., Zhang, H., Hao, T. A hybrid neural network RBERT-C based on pre-trained RoBERTa and CNN for user intent classification. In International Conference on Neural Computing for Advanced Applications, pp. 306-319. Springer, Singapore (2020).
14. Liu, H., Liu, Y., Wong, L. P., Lee, L. K., Hao, T.: A hybrid neural network BERT-cap based on pre-trained language model and capsule network for user intent classification. Complexity 2020, 8858852 (2020).
15. Lee, L. K., Fung, Y. C., Pun, Y. W., Wong, K. K., Yu, M. T. Y., Wu, N. I.: Using a multiplatform chatbot as an online tutor in a university course. In: 2020 International Symposium on Educational Technology (ISET), pp. 53-56. IEEE (2020).
16. Pei, Z., Sun, Z., & Xu, Y. Slang detection and identification. In: Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), pp. 881-889 (2019).
17. Gupta, A., Kumaraguru, P., Castillo, C., Meier, P.: Tweetcred: Real-time credibility assessment of content on Twitter. In: International Conference on Social Informatics 2014, pp. 228-243. Springer, Cham (2014).
18. Lee, J. L., Chen, L., Lam, C., Lau, C. M., Tsui, T. H.: PyCantonese: Cantonese linguistics and NLP in python. In: Proceedings of the 13th Language Resources and Evaluation Conference, pp. 6607-6611. European Language Resources Association (2022).