A new model to automatically detect and filter spam emails

18 March 2022, by Ingrid Fadelli

Credit: Unsplash (Brett Jordan)

Spam emails are undesired messages that are often sent to many random users in bulks. These messages can contain advertisements, but also phishing links or malware. The automatic filtering of emails and the identification of spam messages is highly advantageous, as it can reduce the risk of phishing attacks and make it easier for users to navigate their accounts.

Over the past few years, computer scientists have developed increasingly advanced computational models to automatically detect spam emails. To perform well, however, most of these models need to be trained on large email datasets, which were manually labeled by humans.

Researchers at Sinhgad Institute of Technology Lonavala in India have recently created a new technique for the automatic detection of spam emails. This technique, presented in a paper published in the *International Journal of Intelligent Robotics and Applications,* could help to improve the security of users, while also helping them to skim irrelevant or undesired emails.

"Our model also reduces training speeds and leads to greater efficiency of classification," Vikas Samarthrao Kadam, one of the researchers who carried out the study, told TechXplore. "In contrast with other models, it increases the convergence rate of the spam email detection, achieving better results."

The model developed by Kadam and his colleagues is based on multi-objective feature selection and on an adaptive capsule network, a new and highly promising deep learning technique. In contrast with other previously developed methods, the model was trained on both image and text datasets.

"Our model introduces a new hybrid heuristic algorithm and achieves optimal feature selection, with multi-objective function," Kadam explained. "Our work confirms the promise of new and improved detection models based on deep learning algorithms. The automatic detection of spam emails is necessary due to its simplicity."

The model developed by the researchers is easy to implement and can be trained quickly, over short periods of time. In initial evaluations, Kadam and his colleagues found that it can detect spam emails with greater accuracy than other existing methods.

"Spam detection is essential since it can ensure justice for the sellers and retain the trust of the buyer on the online stores," Kadam said. "In contrast with other methods, it improves the training speed and efficiency of classification. Our model could improve with the quality of life for people who receive large amounts of emails, allowing them to browse through their email smoothly and only use their accounts for their desired purpose."

In the future, the spam filtering technique created by Kadam and his colleagues could be

implemented on a large-scale, improving the security and efficiency of email services. Remarkably, the model can be applied to a wide range of existing services, including Gmail, Yahoo mail and Outlook.

"Almost all researchers present their results based on the accuracy, precision and recall, of their models, but we feel that the time complexity of machine learning models should also be considered as an evaluation metric," Kadam said. "Some researchers show promising results in the process of feature extraction using a bag of words, as they claim that the email header is as important for spam detection as the content of the body. So, deep feature extraction of the header line could also be considered in the future."

So far, the new spam filtering technique devised by this research team achieved very promising results, as it could effectively detect spam emails with a high accuracy. However, Kadam and his colleagues feel that its speed and precision could be improved further in the future.

"The security of spam detection and filtration systems is crucial to achieve better accuracy and reliable results, which can be improved in the future using ensemble learning," Kadam added. "The false positive rate of many models is still higher than required, but it should be reduced to the smallest possible value in future. Real-time spam classification is much needed, as most of the proposed models do not work well with real-time data."

More information: Kadam Vikas Samarthrao et al, A hybrid meta-heuristic-based multi-objective feature selection with adaptive capsule network for automated email spam detection, *International Journal of Intelligent Robotics and Applications* (2022). DOI: 10.1007/s41315-021-00217-9