

Abstract of thesis presented to the Senate of Universiti Malaysia Terengganu in fulfilment of the requirements for the Degree of Doctor of Philosophy

**HYBRID WATER CYCLE OPTIMIZATION ALGORITHM WITH
SIMULATED ANNEALING FOR SPAM EMAIL DETECTION**

GHADA HAMMAD ALRAWASHDEH

AUGUST 2020

Main Supervisor : Rabiei Mamat, Ph.D

Co-Supervisor : Noor Hafhizah Abd Rahim, Ph.D

**Faculty : Faculty of Ocean Engineering Technology and
Informatics**

Spam is referred to unsolicited commercial e-mail from someone trying to give some information that the receiver did not expect. This kind of email usually defined as junk and unwanted. As a filtering step, the spam email filters are implemented in conjunction to reduce this type of e-mails. Unfortunately, new spam email attributes have caused the spam email filter characteristic insufficient and inefficient to handle the large amount of email. This problem is due to the large number of features that the spam classifier needs to evaluate. By the help of feature selection method, the number of features can be reduced. However, the optimal number of features remains a problem and requires further investigation. In this thesis, a new hybrid method has been introduced to make the spam email feature selection more accurate by using the meta-heuristic feature selection optimization approach. The proposed method is based on the hybridization of Water Cycle Algorithm with the Simulated Annealing to optimize the results. This study used a methodology that included groundwork, induction, improvement, assessment, and comparison quality. For the training and validation datasets, cross-validation was performed, and seven datasets were used to evaluate the suggested spam classification. The results show that the water cycle feature selection (WCFS) meta-heuristic was used, as well as three methods of hybridization with Simulated Annealing as a feature selection. In comparison to other feature selection algorithms such as Harmony Search, Genetic Algorithm, and Particle Swarm, the

hybridization interleaved hybridization outperformed them with 96.3 % accuracy, and the effect of using three classifier algorithms revealed that the SVM was the best classifier algorithm with f-measurement 96.3 %. The using interleaved Water Cycle and Simulated Annealing has decreased to more than 50% in number of features. This research contributes to support the Spam Prevention and has an implication practical on the strategy for supporting Spam Prevention, it by control and reduce the rate of spam. The findings could help the companies of antivirus to exert practical effort as well as time on the detection the viruses in the networks internet such reduce the number of big features in the World Wide Web's.

Abstrak tesis yang dikemukakan kepada Senat Universiti Malaysia Terengganu
sebagai memenuhi keperluan untuk Ijazah Doktor Falsafah

**ALGORITMA PENGOPTIMUMAN KITARAN AIR HIBRID DENGAN
PENGUATAN SIMULASI BAGI PENGECEMAN E-MEL SPAM**

GHADA HAMMAD ALRAWASHDEH

OGOS 2020

Penyelia Utama : Rabiei Mamat, Ph.D

Penyelia Bersama : Noor Hafhizah Abd Rahim, Ph.D

Fakulti : Fakulti Teknologi Kejuruteraan Kelautan dan Informatik

Spam dirujuk sebagai e-mel komersil yang tidak diminta, daripada seseorang yang cuba memberikan maklumat yang tidak diharapkan oleh penerima. E-mel jenis ini biasanya ditakrifkan sebagai sampah dan tidak diingini. Sebagai langkah penapisan, penapis e-mel spam dilaksanakan dengan tujuan untuk mengurangkan e-mel jenis ini. Malangnya, atribut e-mel spam yang baru telah menyebabkan ciri penapis e-mel spam tidak mencukupi dan tidak cekap untuk mengendalikan jumlah e-mel yang banyak. Masalah ini disebabkan oleh banyak ciri yang perlu dinilai oleh pengelas spam. Dengan bantuan kaedah pemilihan ciri, bilangan ciri boleh dikurangkan. Walau bagaimanapun, bilangan ciri optimum masih menjadi masalah dan memerlukan siasatan lanjut. Dalam tesis ini, kaedah hibrid baru telah diperkenalkan untuk meningkatkan ketepatan pemilihan ciri e-mel spam dengan menggunakan pendekatan pengoptimuman pilihan meta-heuristik. Kaedah yang dicadangkan adalah berdasarkan gabungan Water Cycle Algorithm dengan Simulasi Annealing untuk mengoptimumkan hasilnya. Metodologi yang digunakan dalam kajian ini yang terdiri daripada asas, induksi, penambahbaikan, penilaian dan kualiti perbandingan. Pengesahan silang digunakan untuk kumpulan latihan dan pengesahan. Disamping itu, tujuh dataset digunakan untuk menguji klasifikasi spam yang dicadangkan. Hasilnya memperlihatkan bahawa meta-heuristik iaitu Water Cycle Feature Selection (WCFS) telah digunakan dengan tiga kaedah gabungan bersama Simulated Annealing sebagai pemilihan ciri yang digunakan.

Perbandingan dengan algoritma-algoritma lain seperti Search Harmony, Algoritma Genetik, dan Partikel Swarm, menunjukkan hasil yang mengagumkan dengan ketepatan sehingga 96.3%. Simulated Annealing telah menurun kepada lebih daripada 50% dalam jumlah ciri. Penyelidikan ini menyumbang untuk menyokong pencegahan Spam dan mempunyai implikasi implikasinya terhadap strategi untuk menyokong pencegahan spam, dengan mengawal dan mengurangkan kadar spam. Penemuan ini dapat membantu syarikat-syarikat antivirus menjalankan usaha praktikal serta masa untuk mengesan virus dalam rangkaian internet seperti mengurangkan jumlah ciri-ciri besar di World Wide Web.