

ในปัจจุบัน มีการคิดค้นวิธีการกำจัดอีเมลขยะขึ้นมาหลากหลายวิธี หนึ่งในวิธีที่นิยมคือการใช้เครื่องเรียนรู้ (Machine Learning) วิทยานิพนธ์นี้ได้นำเสนอ วิธีการกรองอีเมลขยะด้วยทฤษฎีเจเนติกอัลกอริทึม (Genetic Algorithm) ซึ่งกระบวนการหลักๆ ของเจเนติกอัลกอริทึมที่นำมาใช้ประกอบด้วย การคัดเลือก (Selection) การครอสโอเวอร์ (Crossover) และการมิวเตชัน (Mutation) กระบวนการเหล่านี้ถูกนำมาใช้ในการสร้างแม่แบบของอีเมล ซึ่งสืบทอดมาจากอีเมลที่มีอยู่เดิม แม่แบบของอีเมลที่ถูกสร้างขึ้นใหม่จะถูกคัดเลือกเฉพาะแม่แบบที่มีความเหมาะสมมากกว่าแม่แบบอื่นๆ เอาไว้ ตามหลักการของเจเนติกอัลกอริทึม (Fittest of Survival) ทั้งนี้ แม่แบบของอีเมลที่ถูกสร้างขึ้นใหม่ทำให้ตัวกรองได้เรียนรู้รูปแบบของอีเมลใหม่ๆ ส่งผลให้เกิดทางเลือกที่จะนำมาใช้ในการตัดสินใจว่าอีเมลเป็นอีเมลประเภทใดเพิ่มมากขึ้น ซึ่งเมื่อนำตัวกรองอีเมลขยะที่นำเสนอมาทดสอบประสิทธิภาพโดยเปรียบเทียบกับตัวกรองอีเมลขยะโดยประยุกต์ใช้ทฤษฎีเบย์เขียนพบว่ามีประสิทธิภาพที่ดีกว่า โดยมี ค่า Accuracy 87.05% ค่า Recall 88.50% และค่า Precision 86.35%

ABSTRACT

188014

Nowadays, there are several methods to eliminate spam mail. One of popular method is eliminating spam mail by using machine learning. In this paper, we present a spam mail filtering using genetic algorithm. We applied Genetic operator, crossover and mutation to create varieties of mails templates which inherit from old mails. Therefore, it saves time for filter to prepared training set and not need large training set to learn like others method. New mails templates which have more fitness will be selected according to "fittest of survival" of genetic principle. New mail templates are the result of new learning and new choices which improve the spam mail filter efficiency and decided whether the incoming e-mail is spam. In this thesis, we compare the efficiency of filter with Bayesian spam filter. We found that the propose method has more efficient than Bayesian spam filter. With the accuracy 87.05%, recall is 88.50%and precision 86.35%.