

การทำนายเวลาที่ใช้ในการพิมพ์ของผู้ฝึกโดยใช้
วิธีการค่าเฉลี่ยเคกลุ่ม และแบบจำลองต้นไม้เอนไซม์ไฟว์พี

Keystroke Time Interval Prediction

Using K-Means Clustering and Model Tree M5P

กิติชน สุขรุจิ, เด่นดวง ประดับสุวรรณ และ เด่นดวง ประดับสุวรรณ

บทคัดย่อ

งานวิจัยเสนอแนวทางการสร้างแบบจำลองเวลาที่ใช้ในการพิมพ์ของผู้ฝึกพิมพ์ดีดภาษาไทย เนื่องจากการสร้างแบบฝึกใหม่ หรือการจัดตำแหน่งแป้นพิมพ์ใหม่ ซึ่งจำเป็นต้องใช้มนุษย์มาทดลอง แต่ด้วยข้อจำกัดทางด้านเวลา และงบประมาณ ซึ่งเมื่อกลุ่มตัวอย่างผ่านการฝึกไปแล้วจะไม่สามารถใช้กลุ่มตัวอย่างนั้น ๆ มาทดลองได้อีก ผู้วิจัยจึงได้เสนอแนวทางการสร้างตัวแทนชาวยุทธศาสตร์มาทดลองแทนกลุ่มตัวอย่างดังกล่าว เพื่อให้ได้แบบฝึก หรือแป้นพิมพ์ที่เหมาะสมกับผู้ใช้แต่ละคนที่มีทักษะแตกต่างกัน

ในงานวิจัยนี้ได้เสนอวิธีการที่ใช้แบบจำลองต้นไม้ที่ใช้ข้อมูลอินพุตที่ถูกจัดกลุ่มด้วยอัลกอริทึมค่าเฉลี่ยเคกลุ่ม เพื่อสร้างแบบจำลองที่เหมาะสมกับผู้ใช้ โดยเปรียบเทียบผลที่ได้จากวิธีการนำเสนอกับผลการทดลองที่ได้จากแบบจำลองต้นไม้ปกติ โครงข่ายประสาทเทียม และการถดถอยเชิงเส้น โดยวัดจากรากที่สองของความผิดพลาดยกกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ผลการทดลองพบว่าวิธีการที่นำเสนอ มีค่า RMSE เฉลี่ยจากการทดสอบแบบไขว้ข้ามสิบลกลุ่ม (10-fold cross validation) เท่ากับ 3.135 ซึ่งดีกว่าค่าที่ได้จากโครงข่ายประสาทเทียม และการถดถอยเชิงเส้นอย่างมีนัยสำคัญทางสถิติ

Abstract

This research proposes a new method of modeling typing time of each Thai typing learner. Because, in building new typing exercise or re-aligning keyboard layout, testing by human is needed. Due to time, cost and human constraint (that human can test only once), this research proposes a method in which an intelligent agent is trained instead of a user to construct typing lesson or keyboard layout that best fit with each user.

This research proposes an idea of using a tree model which uses clustered data by K-means algorithm to construct a regression model that best fit with each user. The experimental results comparing the ordinary Model Tree, Artificial Neural Network and Linear Regression were measured by Root Mean Square Error (RMSE). The average RMSE obtained from the proposed method is 3.135 centiseconds which is significantly better than other methods run in our experiments using 10-fold cross validation method.