

ชื่อเรื่องการค้นคว้าแบบอิสระ

การรู้จำตัวอักษรภาษาไทยโดยใช้โครงข่ายประสาทเทียม

ผู้เขียน

นายรุจิพันธุ์ โกษารัตน์

ปริญญา

วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

อาจารย์ที่ปรึกษาการค้นคว้าแบบอิสระ

ผู้ช่วยศาสตราจารย์ ดร.เสมอแห สมหอม

บทคัดย่อ

การค้นคว้าแบบอิสระเชิงวิทยานิพนธ์นี้มีวัตถุประสงค์หลักเพื่อศึกษาการนำโครงข่ายประสาทเทียม มาใช้ในการรู้จำอักษรไทย โดยทำการออกแบบและพัฒนาโปรแกรมด้วยภาษาเบสิก แล้วจึงทำการฝึกสอนโครงข่ายประสาทเทียมให้รู้จำอักษรไทย ประกอบด้วย รูปแบบตัวอักษร 2 รูปแบบ คือ อังสนายูพีซี และ คอร์เดียยูพีซี แต่ละรูปแบบมีขนาด 14 16 18 และ 20 พอยต์ โดยมีรูปแบบตัวอักษรแบบปกติ ตัวหนา และ ตัวเอียง

ในขั้นตอนการรู้จำอักษร ได้ทำการแยกอักขระจากรูปภาพโดยอาศัยฮิสโตแกรมในแนวนอนและแนวตั้ง และทำการหาระดับของอักขระในแต่ละบรรทัด โดยได้ทำการแบ่งบรรทัดออกเป็น 4 ระดับ นอกจากนี้ยังนำวิธีวิเคราะห์ทางแบบแฮสคอฟฟีมาใช้ในการแยกสระบนออกจากพยัญชนะในกรณีที่มีการซ้อนทับกัน จากนั้นจึงทำการวิเคราะห์อักขระโดยใช้โครงข่ายประสาทเทียมด้วยค่าน้ำหนักที่ได้จากการฝึกสอน

ผลจากการทดลองพบว่า เมื่อทำการทดลองฝึกสอนโครงข่ายประสาทเทียมด้วยอักขระจำนวน 2 รูปแบบ 3 ประเภท ละ 4 ขนาด และกำหนดขนาดของเมตริกซ์ข้อมูลอินพุท 25x50 พิกเซล สำหรับตัวสระ และ วรรณยุกต์ ขนาดของเมตริกซ์ข้อมูลอินพุท 36x50 พิกเซลสำหรับสำหรับตัวอักษรตัวปกติ และตัวเลข และ ขนาดของเมตริกซ์ข้อมูลอินพุท 62x50 พิกเซล สำหรับตัวอักษรตัวพิเศษ จำนวนนิเวรอนในชั้นซ่อนเท่ากับ 200 โหนด ค่าอัตราการเรียนรู้เท่ากับ 0.3 ค่าความผิดพลาดเฉลี่ยเท่ากับ 0.01 แล้วนำมาทดสอบจากบทความที่มีจำนวนอักขระอยู่ 120 ตัว ผลการทดสอบพบว่า โปรแกรมสามารถวิเคราะห์อักขระได้ถูกต้องเฉลี่ยร้อยละ 97.92 สำหรับตัวธรรมดา ร้อยละ 91.88 สำหรับตัวหนา และร้อยละ 93.13 สำหรับตัวเอียง

Independent Study Title	Thai Language Character Recognition Using Artificial Neural Networks
Author	Mr.Rujipan Kosarat
Degree	Master of Science (Computer Science)
Independent Study Advisor	Asst. Prof. Dr.Samerkae Somhom

ABSTRACT

This independent study has the main purpose to study the application of the Artificial Neural Network (ANN) to recognize the Thai characters. Visual Basic is used in order to design and develop the program, then we train the ANN to recognize Thai sample characters. The samples consist of two different font types; AngsanaUPC, CordiaUPC, with font sizes varies from 14, 16, 18 and 20 points and font format Normal, Bold and Italic.

In the recognition step, a Histogram technique is used to extract characters from pictures vertically and horizontally. Each line is divided into four levels and read the character according to each line level. Besides, in case those characters are overlapped, they are separated by the Hausdorff Distance technique. Finally, the analyze character is with the taken weight from the test.

The result shows that when train the ANN with two fonts, four sizes each and set the input matrix size to 25x50 with to signs, 36x50 to normal and number and 62x50 to special Thai characters at 200 nodes in hidden layer the learning rate is 0.3 with average error of 0.01. When test the program with 120 characters article which has the same font type and size as the samples. It shows that the program can analyze correctly at the average of 97.92 percent at normal font, 91.88 percent at bold font and 93.13 percent at italic font.