บทความวิจัย

Computer Science Department
Faculty of Informatics, Mahasarakham University

การประมวลผลแบบฟอร์มอัตโนมัติ

Automatic Form Processing

อนพัช สนศรี, พงศ์นรา นาคดี, ระพีพร ช่ำชอง
สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม
62011212004@msu.ac.th, 62011212120@msu.ac.th, rapeeporn.c@msu.ac.th

บทคัดย่อ

โครงงานนี้เป็นการวิจัยและพัฒนาการประมวลผลข้อมูลแบบฟอร์มเอกสารที่เป็นข้อมูลภาพไปเป็นข้อมูลตัวอักษร เพื่อศึกษาและพัฒนาเทคนิคการตรวจจับข้อมูลและข้อความบนแบบฟอร์มเอกสาร โดยในงานนี้จะใช้หลักการ Image Processing มาใช้ในขั้นตอน Pre-process เป็นขั้นตอนการเตรียมข้อมูลก่อนเข้าไปประมวลผล และทดสอบอัลกอริทึมที่ใช้ในการรู้จำตัวอักษร โดยใช้การประมวลผลภาพจากการทดลองเพื่อประเมินประสิทธิภาพของ Model สถาปัตยกรรม Convolutional Neural Network (CNN) ร่วมกับ Bi-Directional Long Short Term Memory (BI-LSTM) ในการรู้จำตัวอักษร ซึ่งเลือกใช้ Optimizers 3 ประเภท คือ RMSprop , Adam และ SGD ตามลำดับเพื่อหาผลการเรียนรู้ที่มีประสิทธิภาพที่ดีที่สุด จึงนำสถาปัตยกรรมดังกล่าวมาประยุกต์ใช้กับ Application และ Website เพื่อเป็นการเผยแพร่การใช้งานโปรแกรมประมวลผลการแบบฟอร์มอัตโนมัติ (Automatic Form Processing) ต่อไป

คำสำคัญ: การรู้จำตัวอักษรหลายมือเขียนจากรูปภาพ , Convolutional Neural Network , Bi-Directional Long Short Term Memory, Character Recognition

1.บทนำ

ในปัจจุบันนั้นเทคโนโลยีได้พัฒนาก้าวหน้ามากขึ้นทุกวัน มีเทคโนโลยีที่ตอบสนองความต้องการของมนุษย์มากขึ้นทำให้เราใช้ชีวิตได้สะดวกสบายมากยิ่งขึ้น รวมไปถึงกระบวนการของการแปลงสื่อพิมพ์หรือเขียน เช่นกระดาษ นิตยสาร สัญญา คำร้องขอ หรือข้อมูลอะไรก็ตามที่อยู่ในรูปแบบของเอกสารกระดาษหรือรูปภาพ สามารถทำให้เป็นข้อความ เพื่อใช้ในการคำนวณข้อมูลหรือบันทึกข้อมูล

ซึ่งในปัจจุบันการจัดการเก็บเอกสารที่เป็นสิ่งพิมพ์หรือลายมือเขียนนั้นต้องใช้ทั้งเวลาและทรัพยากรด้านบุคคลในการจัดการกับข้อมูลที่อยู่บนเอกสาร นอกจากนี้อาจมีข้อผิดพลาดต่างๆ ตามมาเช่น การสูญหายของเอกสาร การป้อนข้อมูลในเอกสารผิดพลาด หรือแม้แต่ความไม่สะดวกในการค้นหาข้อมูลที่ต้องการจากเอกสารที่ทำการจัดเก็บเอาไว้

ดังนั้นทางผู้จัดทำจึงได้มุ่งเน้นในการพัฒนาเทคนิคการวิเคราะห์แบบฟอร์ม (Form Analysis) และการรู้จำลายมือ (Handwritten recognition) ตามแบบฟอร์มของเอกสาร เบิก

Computer Science Department
Faculty of Informatics, Mahasarakham University

ค่าลงทะเบียน ของมหาวิทยาลัยมหาสารคาม มาทำการวิเคราะห์แบบฟอร์ม (Form Analysis) และ การรู้จำลายมือ (Handwritten recognition) เมื่อสำเร็จจะสามารถนำข้อมูลที่ได้มาประมวลผล หรือ บันทึกลงในฐานข้อมูลก็ได้ ซึ่งจะทำให้สามารถจัดการเก็บเอกสารดังกล่าวได้ง่ายและรวดเร็วยิ่งขึ้น และสามารถนำไปพัฒนาต่อเพื่อประยุกต์ใช้กับเอกสารในแบบฟอร์มอื่นๆ ได้

2.ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

การตรวจจับเอกสาร ได้ใช้ทฤษฎีการประมวลผลภาพเบื้องต้น มาช่วยจัดการงานทางด้านการเตรียมข้อมูลก่อนเข้าไปเรียนรู้ เพื่อให้การรู้จำตัวข้อความและการสกัดข้อมูลทำงานได้ดียิ่งขึ้น โดยใช้ทฤษฎีประมวลผลเบื้องต้นดังต่อไปนี้

หลังจากอ่านไฟล์ภาพเข้ามาแล้วจะนำภาพไปทำ การแปลงภาพระดับเทา (Gray scale) จากนั้นจะเอาไปทำ การแปลงภาพ GaussianBlur และ เอาไปทำการแปลงภาพ Canny และ ทำการ Contour หาเฉพาะ วัตถุที่เป็น สี่เหลี่ยม แล้วทำการนำตำแหน่งที่ได้จากทำ Contour ไปทำการ Perspective Transformation และทำการ WarpPerspective เพื่อทำให้รูปที่ได้จากการทำ Perspective Transformation การมาตรง และทำการ resize เป็น width = 1357 height = 1920 เพื่อให้ภาพเหมาะกับการ crop ตามตำแหน่งที่เรากำหนดไว้ เพื่อเป็น Input ของกระบวนการรู้จำตัวอักษรต่อไป

3 การรู้จำตัวอักษร จะใช้ Tesseract OCR ย่อมาจาก Optical Character Recognition ซึ่งเป็นกระบวนการของการแปลงสื่อสิ่งพิมพ์ เช่น กระดาษ นิตยสาร สัญญา หรือข้อมูลอะไรก็ตามที่อยู่ในรูปของเอกสารกระดาษ ให้กลายเป็นข้อความให้มีความฉลาดมากขึ้นกว่าการเป็นข้อความธรรมดา โครงสร้างของระบบ OCR ประกอบไปด้วยขั้นตอนการทำงานหลัก 2 ขั้นตอน ได้แก่

3.1 การประมวลผลขั้นต้น (Pre-process) เช่น การปรับแต่งข้อมูล (Normalization) การ กรองข้อมูลแทรกซ้อน (Noise Filtering) การตรวจจับวัตถุ (Object Detection) เป็นต้น

3.2 การรู้จำตัวอักษร (Character Recognition) เช่น วิธีทางโครงข่ายประสาทเทียม เป็นต้น

4. Convolution Neural Network

Convolutional Neural Network (CNN) [3][4] หรือ โครงข่ายประสาทแบบคอนโวลูชัน เป็นโครงข่ายประสาทเทียม แบบทั่วไปโดยจะประกอบไปด้วย ชั้นนำเข้า (Input Layer) ชั้นซ่อน (Hidden Layer)และชั้นผลลัพธ์ (Output Layer) Convolution Neural Network ถูกสร้างขึ้นเพื่อแก้ปัญหานี้โดยทั่วไป CNN ซึ่งจะประกอบไปด้วยชั้นคอนโวลูชันและชั้นพูล

ชั้นคอนโวลูชัน (Convolution Layer) มีชื่อว่าตัวกรองหรือเคอร์เนลซึ่งมีเป้าหมายหลักในการนำฟิลเตอร์ไปใช้กับรูปภาพ การใช้เคอร์เนลกับเมทริกซ์อินพุตช่วยให้สามารถเน้น

Computer Science Department
Faculty of Informatics, Mahasarakham University

ลักษณะของมันได้เช่นขอบวัตถุและอื่น ๆ และนี้คือการดำเนินการ Convolution

ชั้นพูล (Pooling layer) เป็น Layer ที่ทำหน้าที่ในการปรับขนาดและปริมาณของข้อมูลตัวอย่าง(Sample) ให้ลดลงก่อนนำส่งเข้าสู่ Layer ถัดไปเพื่อให้สามารถวิเคราะห์และเก็บรายละเอียดของภาพได้อย่างครบถ้วน โดยในปัจจุบันมีอยู่สองรูปแบบคือ Max pooling, Average pooling

ชั้นฟูลลีคอนเนก (Fully connected layer) เป็น Layer ที่ประกอบด้วยระบบ Multilayer perceptron (MLP) ในการประมวลผลข้อมูลที่ได้มาจาก 2 layer ก่อนหน้านี้เพื่อสังเคราะห์ และทำการแยกแยะรูปภาพออกเป็นหมวดหมู่

5 . Long Short-Term Memory (LSTM)

เครือข่าย Long Short-Term Memory (LSTM) [18] เป็นโครงข่ายประสาทเทียมที่เกิดซ้ำซึ่งแก้ปัญหาการไล่ระดับสีที่หายไป ของ RNN ลักษณะที่โดดเด่นของ LSTM คือ การที่สามารถเลือกได้ว่าข้อมูล ไหนควรที่จะจดจำ ข้อมูลไหนควรกำจัดทิ้ง

6 Gated Recurrent Units (GRU)

Gated Recurrent Units (GRU) [6] มีกลไกภายในที่ เรียกว่า Gate เพื่อควบคุมการไหลของข้อมูลเช่นการจดจำบริบทในช่วงเวลาต่างๆ ติดตามว่าข้อมูลใดใน อดีตสามารถเก็บสิ่งที่ลืมได้ ซึ่งพัฒนาต่อมาจาก LSTM เพื่อลดขั้นตอนการทำงานภายใน Gate และช่วย ให้การประมวลผลเร็วขึ้น

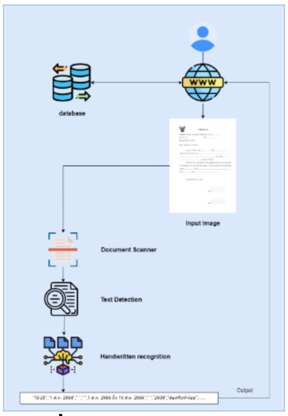
7. การตรวจจับข้อความด้วยการ Contour

การ Contour [13] เป็นเหมือนการหาวัตถุสีดำจากพื้นหลังสีขาว โดยขั้นตอนการ Contour

8. Connection Temporal Classification (CTC)

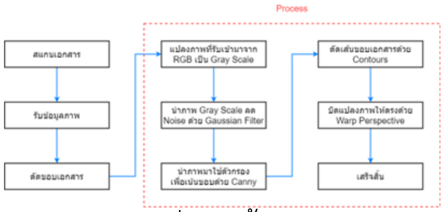
เนื่องจากงานนี้จะเป็นการรู้จำตัวอักษรในการจัดลำดับก่อนหลังว่าอักษรใดมาก่อนและหลังจึงต้องใช้ CTC [21] ในการพิจารณารูปแบบการจัดลำดับของข้อมูลลายมือเขียนที่มีความต่อเนื่องกัน

3.ทฤษฎีและงานวิจัยที่เกี่ยวข้อง



ภาพประกอบที่ 3.1 กรอบการดำเนินงานของระบบ

ส่วนที่ 1: การสแกนหาเอกสาร (Document Scanner)



ภาพประกอบที่ 3.2 ขั้นตอนในการทำ Document Scanner

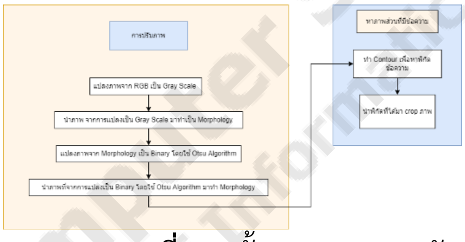
Computer Science Department
Faculty of Informatics, Mahasarakham University

ขั้นตอนนี้การทำ Document Scanner รับข้อมูลภาพเข้ามา (Input) โดยข้อมูลภาพที่รับเข้ามาจะเป็นภาพ RGB แปลง ระดับเทา (Gray Scale) มาแล้วทำการลดสัญญาณรบกวน (Noise) ด้วย (Gaussian Filter) แล้วทำการนำภาพใช้ตัวกรองเพื่อค้นหาหรือเน้นส่วนขอบด้วยตัวกรอง Canny หลังจากใช้ตัวกรอง Canny ทำการวาดเส้นเค้าโครงลงบนภาพซึ่งมี Draw Contours และ Find Contours แล้วทำการบิดแปลงภาพให้ตรงด้วย Warp Perspective

ส่วนที่ 2: การตัดภาพ (Crop)

จะทำการตัดเอาภาพส่วนที่เราต้องการเพื่อนำเข้าไปสู่กระบวนการการ การตรวจจับข้อความ (Text Detection) ในขั้นตอนต่อไป

ส่วนที่ 3: การตรวจจับข้อความ (Text Detection)



ภาพประกอบที่ 3.3 ขั้นตอนการตรวจจับข้อความ

ขั้นตอนนี้คือขั้นตอนการเตรียมภาพก่อนเข้าไปสู่การรู้จำตัวอักษรเมื่อแปลงภาพมาให้อยู่ในระดับ Gray Scale แล้วจะนำภาพที่ได้มาทำ MORPH_GRADIENT เพื่อเอาเส้นขอบเค้าโครงของภาพ และนำภาพที่ได้จากการทำ MORPH_GRADIENT แล้วนำไปแปลงเป็นภาพสองระดับด้วย Thresholding โดยใช้ Otsu Algorithm จากนั้นนำภาพขาวดำไปทำ. MORPH_CLOSE เพื่อทำให้เส้นขอบเค้าโครงของภาพพองขึ้น และจึงนำภาพนั้นไปหาพิกัดของบรรทัดและนำไป Crop เพื่อเป็น Input ของกระบวนการการรู้จำลายมือ (Handwritten recognition)

3.2 การสแกนหาเอกสาร (Document Scanner)

1. การรับข้อมูลภาพเข้ามา (Input Image)

โดยข้อมูลภาพที่รับเข้ามาจะอยู่ในรูปแบบไฟล์ภาพ PNG, JPG เป็นต้น

2. การแปลงภาพสี RGB ไปเป็นภาพระดับเทา (Gray Scale)

ภาพระดับสีเทา (Gray Scale) คือภาพขาว - ดำ - เทา โดยจะมีระดับความเข้มจากการแปลงสีเท่าคือ 0–255 (8-bit)

3. การลดสัญญาณรบกวน (Noise) ภาพระดับเทาด้วยตัวกรองเกาส์เซียน (Gaussian Filter)

เป็นตัวกรองสัญญาณรบกวนที่มีคุณลักษณะคล้ายระฆังคว่ำ ใช้สำหรับลดสัญญาณรบกวน (Noise) และลบความคมชัดของรูปภาพ

4. การทำตรวจจับขอบ (Edge Detection) โดยใช้ตัวกรอง Canny

เป็นการหาเส้นรอบวัตถุที่อยู่ในภาพ หาเส้นรอบวัตถุเนื่องจากขอบภาพเกิดจากความแตกต่างของความเข้มแสงจากจุดหนึ่งไปยังอีกจุดหนึ่งหากความต่างนี้มีค่ามากจะทำให้เห็นขอบภาพได้อย่างชัดเจน

5. การหาเส้นเค้าโครง (Contours) ของเอกสาร

Computer Science Department
Faculty of Informatics, Mahasarakham University

6. การบิดแปลงภาพให้ตรงด้วย Warp Perspective

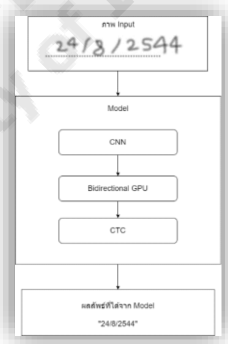
3.3 การตัดภาพ (Crop)

เป็นการหาค่าความกว้างและความสูงเพื่อระบุตำแหน่งที่จะเอามาตัด crop รูปภาพให้เหลือเฉพาะส่วนที่ต้องการ

ภาพประกอบที่ 3.35 ตัวอย่างการระบุตำแหน่งที่ต้องการตัด

ภาพประกอบที่ 3.36 ผลลัพธ์การตัด crop

3.4 การรู้จำลายมือ (Handwritten recognition)



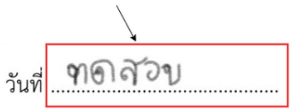
ภาพประกอบที่ 3.67 ขั้นตอนการทำงานของการรู้จำลายมือ

เมื่อผ่านกระบวนการ Pre-Processing เสร็จแล้วผลลัพธ์ที่ได้จากขั้นตอน Pre-Processing จะถูกส่งเข้าสู่กระบวนการรู้จำตัวลายมือ

3.5 ชุดข้อมูลที่ใช้ในการทดลอง

ชุดข้อมูลที่ใช้ในการสร้างโมเดลรู้จำมากกว่า 200 ภาพจะแบ่งออกเป็นชุดสอน 60% ชุดวาลิเดท 20% และชุดทดสอบ 20% ซึ่งนำมาตัดเป็นข้อมูล 1 บรรทัด

ภาพประกอบที่ 3.82 ตัวอย่างชุดข้อมูลเอกสาร เบิกค่าลงทะเบียน (ชุดข้อมูล 1 ภาพ)



ภาพประกอบที่ 3.83 ตัวอย่างชุดข้อมูลเอกสาร เบิกค่าลงทะเบียน (ชุดข้อมูล 1 บรรทัด)

3.6 การวัดประสิทธิภาพ

การประเมินความถูกต้องของการรู้จำตัวอักษร โดยใช้ Levenshtein Edit Distance ซึ่งเป็นขั้นตอนการวัดเพื่อหาค่าความแตกต่างของข้อความทั้งสองชุดระหว่างชุดแรกที่เป็นผลเฉลย และชุดที่สองที่เป็นชุดผลการทำนาย โดย

จะนับจำนวนครั้งของการแทรก การตัด และการแทนที่

Computer Science Department
Faculty of Informatics, Mahasarakham University

4. ผลการทดลอง

4.1 ผลการทดลองขั้นตอนการ Pre-process

ในขั้นตอนนี้ จะแสดงผลการทดลองการเตรียมภาพเพื่อนำไปประมวลผล โดยประกอบไปด้วย 6 ขั้นตอน

1.ภาพ input

2.การทำ gray scale

3.แปลงภาพ GaussianBlur

4.แปลงภาพ Canny

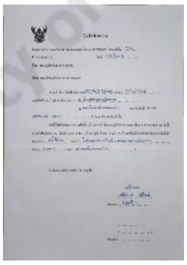
5.ทำ Contours

6. ทำการ crop ตามพิกันที่ Contours

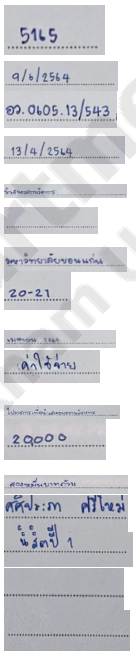
4.2 การตัดภาพส่วนที่มีข้อความ

ในขั้นตอนนี้จะทำการตัดภาพให้เหลือเฉพาะส่วนที่มีข้อความอยู่เท่านั้น โดยจะยกตัวอย่างผลลัพธ์ดังตารางต่อไปนี้

ตารางที่ 4.5 ตัวอย่างผลลัพธ์ของการตัดภาพส่วนที่มีข้อความ

ขั้นตอนที่	ภาพ	การทำงาน
1		ภาพ input ที่ได้จากการ pre-process

ตารางที่ 4.6 ตัวอย่างผลลัพธ์ภาพที่ได้จากการ Crop

ขั้นตอนที่	ภาพ	การทำงาน
2		ภาพที่ได้จากการCrop พิกัดที่ระบุไว้

4.3 การเปรียบเทียบประสิทธิภาพการรู้จำโดยการปรับค่า Optimizers

4.3.1 การประเมินประสิทธิภาพ Model

ในการประมวลผลแบบฟอร์มอัตโนมัตินี้ ได้ใช้สถาปัตยกรรม CNN ร่วมกับ RNN โดยได้ทำการเลือกใช้ Optimizers 3 ประเภท ได้แก่ RMSprop , Adam และ RMSPROP ตามลำดับเพื่อหาผลการเรียนรู้ที่มีประสิทธิภาพที่ดีที่สุด โดยในการทดลองนี้ ได้ใช้ข้อมูลในการเรียนรู้ทั้งหมด 7,222 ภาพ และทำการแบ่งชุดข้อมูลออกเป็น 3 ชุดได้แก่ ชุดเรียนรู้ 60% ชุดวาลิเดชัน 20% และชุดทดสอบ 20% ใช้จำนวน

Computer Science Department
Faculty of Informatics, Mahasarakham University

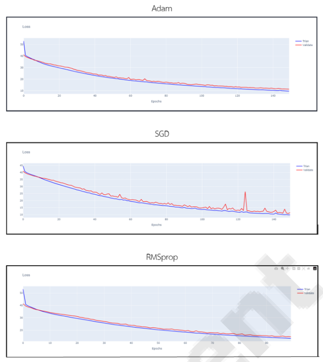
รอบ (epoch) ในการรู้จำทั้งหมด 100 รอบ และในชุดเรียนรู้ทำการ Augmentation รูปภาพเพิ่มเติม 7,222 ภาพ ทำให้มีภาพในชุดเรียนรู้ทั้งสิ้น 14,444 ภาพ

ตารางที่ 4.7 รูปแบบ Model ที่ทำการทดลองทั้ง 2 Model

Model
Input Image 150 x 32
CNN 2 Layer (32,64)
Bidirectional LSTM 2 Layer (256,128)
SoftMax Activation ReLU Optimizer
CTC
Output

ตารางที่ 4.8 ผลการประเมินประสิทธิภาพค่า CER ที่ได้จากการเรียนรู้ Optimizer ทั้ง 3 ประเภท

Optimizer	CER (%)
Adam	8.107058333
SGD	15.56849167
RMSPROP	12.95589167

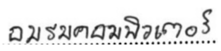


ภาพประกอบที่ 4.1 กราฟโมเดล 3 โมเดล Adam, SGD, RMSprop

จากกราฟผลการเรียนรู้ข้างต้น จะเห็นว่าการเรียนรู้โดยใช้ Optimizer RMSprop และ Adam พบว่าค่า Loss มีการลดลงอย่างรวดเร็ว ส่วน Optimizer SGD นั้นพบว่าค่า Loss มีอัตราการลดลงค่อนข้างช้ากว่า 2 วิธี

จากผลการเรียนรู้ด้วย Optimizer ทั้ง 3 ประเภท จากการเรียนรู้เป็นจำนวน 150 รอบ พบว่า SDG นั้นมีค่า Character Error Rate (CER) ต่ำสุด และ Adam มีค่าสูงสุด

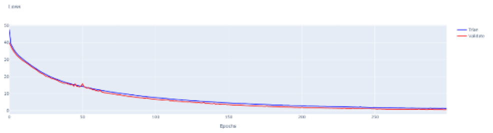
ตารางที่ 4.9 ตัวอย่างผลลัพธ์การทำนายผลจาก Optimizer ทั้ง 3 ประเภท ภาพที่ 1

ภาพต้นฉบับ		CER (%)
ผลเฉลย	อบรมคอมพิวเตอร์	
Optimizer Adam	'อบรคอมพิวเตอร์'	6.6667
Optimizer SGD	'อบมคอพิวเตอร์'	13.3333
Optimizer RMSPROP	'อบรมอมพิวเตอร์'	6.6667

Computer Science Department
Faculty of Informatics, Mahasarakham University

4.4 ผลลัพธ์ประสิทธิภาพการรู้จำจาก Model Optimizer

Optimizer Adam ให้ประสิทธิภาพที่ดีที่สุดจึงเลือกใช้วิธีดังกล่าว และได้ใช้ข้อมูลในการเรียนรู้ทั้งหมด 21,695 ด้วยชุดข้อมูลสำหรับเรียนรู้ 60% ชุดวาลิเดชัน 20% และชุดทดสอบ 20% จากข้อมูลทั้งหมด แสดงผลการฝึกทั้ง 300 รอบดังตางรางต่อไปนี้



ภาพประกอบที่ 4.2 ภาพกราฟแสดงค่า Loss ที่ได้จาก Model

ตารางที่ 4.21 ตัวอย่างผลลัพธ์การทำนายกับเอกสารจริง ภาพที่ 1

<table>
<tr><td rowspan="3">ใบที่ 1
 </td><td colspan="2">ข้อความรู้จำถูกต้อง</td><td rowspan="2">CER (%)</td></tr>
<tr><td>จำนวนข้อความ</td><td>ความถูกต้อง (%)</td></tr>
<tr><td>2</td><td>11.76</td><td>88.5455</td></tr>
<tr><td>ผลเฉลย</td><td colspan="3">‘5123’ ‘9/6/2564’ ‘อว. 0605.13/123’ ‘12/4/2564’ ‘อบรมคอมพิวเตอร์’ ‘’ ‘ม. ขอนแก่น’ ‘15-20’ ‘พฤษภาคม2564’ ‘อุนหนุน’ ‘อบรม’ ‘20,000’ ‘สองหมื่นบาทถ้วน’ ‘นางสาวพิรดา ถาวร’ ‘นิสิตชั้นปีที่2’ ‘’ ‘’</td></tr>
<tr><td>ผลที่ถูกทำนาย</td><td colspan="3">'24' '29มคม/2564' 'อว. 060513/54' '1ษาค0' 'แปแดมึ๋น้ายน' 'อีชำชฉะ' '29ุ๊กนายน 2560' '15-20"ณิุ้าทา' 'อัหนุน' 'แ็งหดิจ้าม2ม5มม' '20,000' 'แอ์งโน้้านา2ว.' 'พฤษภาคม2565' 'ออุ๋นหาคนน' '2ปางพันบนาทถ้' 'ปนบชะ'</td></tr>
</table>

4.5 เปรียบเทียบประสิทธิภาพการรู้จำใบเสร็จจาก Tesseract

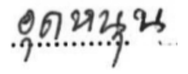
ในการทดลองนี้ได้ทำการเปรียบเทียบการรู้จำเอกสารที่มารายมือเขียนจากวิธีที่พัฒนากับการใช้ Tesseract โดยได้ผลการทดลอง ดังนี้

ตารางที่ 4.31 ตัวอย่างการทำนายผลระหว่าง Model ที่ทำมาเทียบกับ Tesseract ภาพที่ 1

ภาพต้นฉบับ		CER (%)
ผลเฉลย	อาจารย์	
Model Optimizer Adam	อาจารย๋	14.2857
Tesseract	(ยาจาา๑	71.4285

Computer Science Department
Faculty of Informatics, Mahasarakham University

ตารางที่ 4.32 ตัวอย่างการทำนายผลระหว่าง Model ที่ทำมาเทียบกับ Tesseract ภาพที่ 2

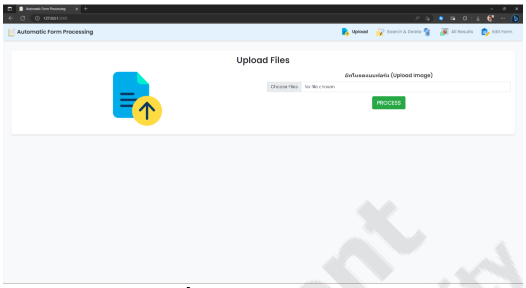
ภาพต้นฉบับ		CER (%)
ผลเฉลย	อุดหนุน	
Model Optimizer Adam	อุดหนุน	0
Tesseract	จดหษุพ	57.1428

จากการทดลองดังกล่าว ได้ข้อสรุปว่า Model โดย Tesseract มีค่า CER เฉลี่ยทั้งอยู่ที่ 70.2688% ในขณะที่การใช้ Model Optimizers Adam มีค่า CER อยู่ที่ 25.0184%

4.6 ตัวอย่างการทำนายผลบน Mobile Application

ภาพประกอบที่ 4.3 ตัวอย่างก่อนทำนาย Website บน Mobile

4.7 ตัวอย่างการทำนายผลบน Website



ภาพประกอบที่ 4.8 ตัวอย่างก่อนทำนายบน Website

4.8 สรุปและวิเคราะห์ผลการทดลอง

จากการทดลองในการหาตัว Optimizers ที่ดีที่สุดโดยทำการทดลองจากภาพจำนวน 14,444 ภาพ พบว่า ตัว Optimizers Adam มีค่า CER คิดเป็น 8.1070% ซึ่งต่ำที่สุด เมื่อเทียบกับ ตัว Optimizers อื่น จาก ตารางการทดลองตารางที่ 4.8 ผลการประเมินประสิทธิภาพค่า CER ที่ได้จากการเรียนรู้ Optimizer ทั้ง 3 ประเภท การเปรียบเทียบประสิทธิภาพการรู้จำโดยการปรับค่า Optimizers จึงทำให้สรุปได้ว่า Optimizer Adam มีประสิทธิภาพที่ดีกว่า Optimizer อื่น

ในส่วนของการเปรียบเทียบประสิทธิภาพการรู้จำเอกสารพบว่า การใช้ Model Optimizer Adam จะทำให้ได้ผลลัพธ์ดีกว่าการใช้ Tesseract ซึ่งมีค่า CER อยู่ที่ 70.2688% ดังนั้น ในงานวิจัยนี้จึงเลือกใช้ Optimizer Adam และทำการทดลองกับชุดข้อมูล 204 ภาพ ตัดเป็นข้อความได้ทั้งหมด 3,611ภาพ และนำไปทำ Augmentation รูปภาพเพิ่มเติม 18,055 ภาพ ทำให้มีภาพในชุดเรียนรู้ทั้งสิ้น 21,666 ภาพ ใช้ชุดข้อมูล

Computer Science Department
Faculty of Informatics, Mahasarakham University

สำหรับเรียนรู้ 60% ชุดวาลิเดชัน 20% และชุดทดสอบ 20% ฝึกทั้ง 300 รอบ โดยมีค่า Character Error Rate (CER) เป็น 25.0184% มาประยุกต์ใช้กับ Application และ Website เพื่อเป็นการเผยแพร่การใช้งานโปรแกรมการประมวลผลแบบฟอร์มอัตโนมัติ (Automatic Form Processing) ต่อไป

5. สรุปและอภิปรายผลการทดลอง

โครงงานปริญญานิพนธ์ฉบับนี้นำเสนอการประมวลผลแบบฟอร์มอัตโนมัติ (Automatic Form Processing) โดยใช้การประมวลผลภาพเบื้องต้นและการเรียนรู้เชิงลึกมาใช้งานนี้ ซึ่งสามารถใช้งานการประมวลผลแบบฟอร์มได้ผ่าน Website Application

จากการทดลองวัดประสิทธิภาพการรู้จำตัวอักษรจากฐานข้อมูลรูปภาพรวมกันทั้งสิ้น 21,666 ภาพ โดยใช้ Optimizer Adam ซึ่งมีประสิทธิภาพในการทำนายผลสูงสุด โดยมีค่า Character Error Rate (CER) ต่ำสุดอยู่ที่ 25.0184%

5.1 ปัญหา และอุปสรรคในการดำเนินงาน

1. เนื่องจาก Algorithm ที่ใช้ในการรู้จำมีความซับซ้อนมากจำเป็นต้องใช้ทรัพยากรในการประมวลผลสูงและใช้ระยะเวลานาน

2. ชุดข้อมูลเป็นภาพที่ถ่ายหรือสแกนจากมือถือ มีสัญญาณรบกวนไม่ว่าจะเป็นแสง ขนาดของภาพ คุณภาพของรูปถ่าย องศาของภาพถ่าย ผลลัพธ์ที่ได้จากการประมวลผลแบบฟอร์มจึงมีความคลาดเคลื่อน

3. Algorithm การประมวลผลแบบฟอร์มอัตโนมัติ จำเป็นต้องมีการทำนายผลเฉลยด้วยตัวเองจึงใช้ระยะเวลานานในการสร้างผลเฉลย

5.2 ข้อเสนอแนะ

1. ควรพัฒนาเพิ่มเติมในส่วนของข้อมูลเอกสารที่ใช้ในการเทรนให้มีหลากหลายมากขึ้น

2. ควรปรับปรุงการกำจัดสัญญาณรบกวน แสงเงา การตัดภาพ และการปรับปรุงคุณภาพของรูปภาพเพื่อให้ได้คุณภาพที่ดีขึ้น

เอกสารอ้างอิง

1. “การรู้จำตัวอักษรภาษาไทยที่เขียนด้วยลายมือในแบบฟอร์ม (Thai Character Recognition of Handwritten in Forms)”. https://shorturl.asia/DW70B (สืบค้น ต.ค 26, 2021).

2. “การรู้จำ OCR ด้วย Python และ Tesseract”. https://ichi.pro/th/phaen-ngan-pheux-kar-ru-ca-xakkhra-dwy-saeng-dwy-python-laea-tesseract-128751611475191 (สืบค้น ต.ค 26, 2021).

3. “Thai Written Net : Thai Hand Written Script Recognition Using Deep Neural Networks”. http://azjhpc.com/issue5/doi.org.10.32010.26166127.2020.3.1.75.93.pdf (สืบค้น ต.ค 26, 2021).

4. “การประมวลผลลายมือเขียนเป็นตัวพิมพ์อัตโนมัติ เวอร์ชั่น 2 (Automatic Handwritten Recognition; Auto HWR v.2)

Computer Science Department
Faculty of Informatics, Mahasarakham University

". http://digital.csmsu.net:8080/library/handle/123456789/105 (สืบค้น ต.ค 26, 2021).

5. "Image Processing". https://medium.com/@loved_pastel_alpaca_906/image-processing-445f856d62cc (สืบค้น ต.ค. 26, 2021).

6. "R. Fisher, S. Perkins, A.Walker and E. Wolfart. – Thresholding". https://www.dsi.unive.it/~atorsell/Hipr.pdf (สืบค้น ต.ค 26, 2021).

7. "Natthawat Phongchit - Convolution Neural Network (CNN) คืออะไร". https://medium.com/@natthawatphongchit/มาลองดูวิธีการคิดของ-cnn-กัน-e3f5d73eebaa (สืบค้น ต.ค 26, 2021).

8. "Natthawat Phongchit - ย้อนรอย Object Detection และเจาะรึก Retina Net". https://medium.com/@natthawatphongchit/ย้อนรอย-object-detection-และเจาะลึก-retinanet-ea1407ad7325 (สืบค้น ต.ค 26, 2021).

9. "Arthit Suriyawongkul - ทดลอง Tesseract 4.0alpha กับภาษาไทย". https://medium.com/@bact/ทดลอง-tesseract-4-0alpha-กับภาษาไทย-8248a73c5ae5 (สืบค้น ต.ค 26, 2021).

10. "Phyblas - opencv-python บทที่ 4 การจัดสี". https://phyblas.hinaboshi.com/oshi04 (สืบค้น ต.ค 26, 2021).

11. "Phyblas - opencv-python บทที่ 7 การหมุนหรือบิดแปลงภาพ". https://phyblas.hinaboshi.com/oshi07 (สืบค้น ต.ค 26, 2021).

12. "Phyblas - opencv-python บทที่ 8 ตัวกรองคอนโวลูชันและการทำภาพเบลอ". https://phyblas.hinaboshi.com/oshi08 (สืบค้น ต.ค 26, 2021).

13. "Phyblas - opencv-python บทที่ 9 การใช้ตัวกรองเพื่อค้นหาหรือเน้นส่วนขอบ". https://phyblas.hinaboshi.com/oshi09 (สืบค้น ต.ค 26, 2021).

14. "Phyblas - opencv-python บทที่ 12 การกร่อน พองตัว และแปลงสัณฐาน". https://phyblas.hinaboshi.com/oshi12 (สืบค้น ต.ค 26, 2021).

15. "Phyblas - opencv-python บทที่ 13 การหาเส้นเค้าโครง". https://phyblas.hinaboshi.com/oshi13 (สืบค้น ต.ค 26, 2021).

16. "EGBE443 Image Processing in Medicine, BME, MU - Gradient and Edge Detection". https://imageprocessinggr3.wordpress.com (สืบค้น ต.ค 26, 2021).

17. "Mattro Tortoli – Convolution Network: theory and code". https://matteotor92.medium.com/convolutional-neural-network-theory-and-code-20bbe066fa48 (สืบค้น ต.ค 26, 2021).

Computer Science Department
Faculty of Informatics, Mahasarakham University

18. “Sequence Modeling With CTC”. https://distill.pub/2017/ctc (สืบค้น พ.ย 12, 2021).

19. “Optical character recognition”. https://en.wikipedia.org/wiki/Optical_character_recognition (สืบค้น พ.ย 12, 2021).

20. “Long Short Term Memory and Gated Recurrent Unit’s Explained — ELI5 Way”. https://towardsdatascience.com/long-short-term-memory-and-gated-recurrent-units-explained-eli5-way-eff3d44f50dd (สืบค้น พ.ย 12, 2021).

21. “How to easily do Handwriting Recognition using Machine Learning”. https://nanonets.com/blog/handwritten-character-recognition (สืบค้น พ.ย 12, 2021).

22. “การใช้ Tesseract ทำ OCR ภาษาไทย”. https://medium.com/super-ai-engineer/การใช้-tesseract-ทำ-ocr-ภาษาไทย-94e5c5863ae5 (สืบค้น พ.ย 12, 2021).

23. “Automatic summarisation of legal documents”. https://researchers.mq.edu.au/en/publications/automatic-summarisation-of-legal-documents (สืบค้น พ.ย 12, 2021).

24. “Urdu word segmentation”. https://aclanthology.org/N10-1077.pdf (สืบค้น พ.ย 12, 2021).

25. “Effectiveness of Different Similarity Measures for Text Classification and Clustering”. https://www.semanticscholar.org/paper/Effectiveness-of-Different-Similarity-Measures-for-Maher-Joshi/1e93d3191b281f2cbea39340f709488796978595 (สืบค้น พ.ย 12, 2021).

26. “Efficient Hyperparameter Tuning with Grid Search for Text Categorization using kNN Approach with BM 25 Similarity”. https://www.degruyter.com/document/doi/10.1515/comp-2019-0011/html?lang=en (สืบค้น พ.ย 12, 2021).

27. “Fuzzy Bag-of-Words Model for Document Representation”. https://ieeexplore.ieee.org/document/7891009 (สืบค้น พ.ย 12, 2021).

28. “Clustering More than Two Million Biomedical Publications: Comparing the Accuracies of Nine Text-Based Similarity Approaches”. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0018029 (สืบค้น พ.ย 12, 2021).

29. “The Role of Text Pre-processing in Sentiment Analysis”. https://www.sciencedirect.com/science/article/pii/S1877050913001385 (สืบค้นพ.ย 12, 2021).

Computer Science Department
Faculty of Informatics, Mahasarakham University

30. “Sentence Classification Experiments for Legal Text Summari sation”. http://jurix.nl/pdf/j04-05.pdf (สืบค้น พ.ย 12, 2021).

Computer Science Department
Faculty of Informatics, Mahasarakham University