การหาค่าเหมาะสมที่สุดของการเรียนรู้เชิงลึกโดยใช้อัลกอริทึมเชิงวิวัฒนาการ

โดย
นายราชฤทธา รัตนวรกานต์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
ภาควิชาวิศวกรรมไฟฟ้า
บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร
ปีการศึกษา 2562
ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

การหาค่าเหมาะสมที่สุดของการเรียนรู้เชิงลึกโดยใช้อัลกอริทึมเชิงวิวัฒนาการ

โดย
นายราชฤทธา รัตนวรกานต์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
ภาควิชาวิศวกรรมไฟฟ้า
บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร
ปีการศึกษา 2562
ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

HYPER-PARAMETER OPTIMIZATION FOR DEEP NEURAL NETWORK USING AN ISLAND-BASED GENETIC ALGORITHM

By

MR. Rashrita RATTANAVORAGANT

A Thesis Submitted in Partial Fulfillment of the Requirements

for Master of Engineering (ELECTRICAL AND COMPUTER ENGINEERING)

Department of ELECTRICAL ENGINEERING

Graduate School, Silpakorn University

Academic Year 2019

Copyright of Graduate School, Silpakorn University

หัวข้อ	การหาค่าเหมาะสมที่สุดของการเรียนรู้เชิงลึกโดยใช้อัลกอริทึมเชิงวิวัฒนาการ
โดย	ราชฤทธา รัตนวรกานต์
สาขาวิชา	วิศวกรรมไฟฟ้าและคอมพิวเตอร์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
อาจารย์ที่ปรึกษาหลัก	ผู้ช่วยศาสตราจารย์ ดร. ยุทธนา เจวจินดา

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร ได้รับพิจารณาอนุมัติให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.จุไรรัตน์ นันทานิช)

พิจารณาเห็นชอบโดย

.....ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ยรรยงค์ พันธ์สวัสดิ์)

.....อาจารย์ที่ปรึกษาหลัก
(ผู้ช่วยศาสตราจารย์ ดร.ยุทธนา เจวจินดา)

.....ผู้ทรงคุณวุฒิภายนอก
(ดร.กนก พานทอง)

60407210 : วิศวกรรมไฟฟ้าและคอมพิวเตอร์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
คำสำคัญ : การเรียนรู้เชิงลึก, อัลกอริทึมเชิงวิวัฒนาการ

นาย ราชฤทธา รัตนวรกานต์: การหาค่าเหมาะสมที่สุดของการเรียนรู้เชิงลึกโดยใช้อัลกอริทึมเชิงวิวัฒนาการ อาจารย์ที่ปรึกษาวิทยานิพนธ์ : ผู้ช่วยศาสตราจารย์ ดร. ยุทธนา เจวจินดา

งานวิจัยฉบับนี้ได้เสนอวิธีการหาค่าเหมาะสมที่สุด(hyper-parameters)ของการเรียนรู้เชิงลึกโดยอัลกอริทึมเชิงวิวัฒนาการ โดยอัลกอริทึมที่เลือกใช้ในงานวิจัยฉบับนี้เป็นอัลกอริทึมที่มีชื่อว่า Island-based genetic algorithm หรือ ISLGA อัลกอริทึมนี่เป็นอัลกอริทึมที่มีลักษณะเด่นที่แตกต่างอัลกอริทึมเชิงวิวัฒนาการทั่วไปคือ ในอัลกอริทึมนี่จะทำการแบ่งประชากรทั้งหมดเป็นกลุ่มย่อย หลายกลุ่ม และเพิ่มกลไกการแลกเปลี่ยนประชากรระหว่างเกาะ จากลักษณะดังกล่าวนี้ทำให้คำตอบของปัญหา หรือในงานวิจัยนี้ ค่า hyper-parameters ที่หาได้ไม่ลู่เข้าสู่ค่าใดค่าหนึ่งเร็วเกินไป ทำให้ค่าที่ได้จากอัลกอริทึมนั้นเป็นค่า hyper-parameter ที่เหมาะสมที่สุดสำหรับการเรียนรู้เชิงลึก

วิธีการที่ถูกเสนอประกอบด้วยสองขั้นตอนหลัก ได้แก่ การหาค่า hyper-parameter และการฝึกสอนการเรียนรู้เชิงลึก เพื่อเป็นการนำเสนอประสิทธิภาพของอัลกอริทึม ผู้วิจัยได้ทำการทดลองโดยการนำอัลกอริทึมนี้มาใช้ในการหาค่า hyper-parameter ของการเรียนรู้เชิงลึกชนิด Deep convolutional neural network (DCNN) ซึ่งการเรียนรู้เชิงลึกที่ใช้ในการทดลองมีต้นแบบมาจากการเรียนรู้เชิงลึกที่มีชื่อว่า Alexnet ผู้วิจัยได้ทำการเข้ารหัสโครงสร้างและ hyper-parameter ในรูปของโครโมโซมเพื่อนำไปใช้ในอัลกอริทึม ISLGA .ในการทดลองผู้วิจัยได้ทำการทดสอบค่า hyper-parameter ที่หาได้จากวิธีการที่นำเสนอโดยการให้ทำการแยกแยะภาพจากฐานข้อมูล CIFAR-10 จากการทดลองพบว่าอัลกอริทึมนี่สามารถหาค่า hyper-parameter ของการเรียนรู้เชิงลึกได้เหมาะสมและมีประสิทธิภาพ

60407210 : Major (ELECTRICAL AND COMPUTER ENGINEERING)
Keyword : Deep convolutional neural network, Genetic algorithm, Island-based genetic algorithm

MR. RASHRITA RATTANAVORAGANT : HYPER-PARAMETER OPTIMIZATION FOR DEEP NEURALNETWORK USING AN ISLAND-BASED GENETIC ALGORITHM THESIS ADVISOR : ASSISTANT PROFESSOR DR. YUTANA JEWAJINDA

In this research, we propose the method using the island-based genetic algorithm (ISLGA) to automatically optimize the hyper-parameter of deep neural networks (DNNs). The ISLGA overall process is the same as a common genetic algorithm but possesses unique characteristic and method which is the separation of the population into a small group and force the part of the population to migrate in every set period of generation. With the mentioned characteristic, the chance of early convolution of the answer is significantly lower which leads to better hyper-parameters selection. The proposed method consists of two steps: hyper-parameter searching and complete DNNs training. To demonstrate our method, we apply our algorithm to find the hyper-parameter for Deep convolutional neural network (DCNN) which design is based on the DCNN call Alexnet. We encode the structure and hyper-parameters of chosen DCNN as a chromosome and then apply an ISLGA to optimize the structure and hyper-parameters. To evaluate the performance of the proposed method, we implemented an image classification task using CIFAR-10 dataset. The experimental result shows that the proposed method can efficiently search for the appropriated hyper-parameters of the DCNNs.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ด้วยความกรุณาจากผู้ช่วยศาสตราจารย์ ดร. ยุทธนา เจวจินดา ซึ่งเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ที่ให้คำแนะนำและความช่วยเหลือแก่ผู้วิจัยตลอดการกระบวนการวิจัย รวมทั้ง อาจารย์ ดร. ยรรยงค์ พันธ์สวัสดิ์ ประธานกรรมการสอบวิทยานิพนธ์ และผู้ช่วยศาสตราจารย์ทวีศักดิ์ สมานชื่น ผู้ทรงคุณวุฒิที่กรุณาให้คำปรึกษาและข้อเสนอแนะที่เป็นประโยชน์แก่ผู้วิจัย ทำให้วิทยานิพนธ์เล่มนี้ มีความถูกต้องและสมบูรณ์ ยิ่งขึ้น

ผู้วิจัยขอกราบขอบพระคุณในความกรุณาของอาจารย์ทุกท่านเป็นอย่างสูง ผู้วิจัยขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาวิศวกรรมไฟฟ้าทุกท่าน ที่ให้ความรู้, การสนับสนุนไม่ว่าจะเป็นด้านทุนสำหรับการศึกษาเล่าเรียน อุปกรณ์ต่างๆสำหรับงานวิจัย และประสบการณ์อันมีค่าแก่ผู้วิจัย ท้ายที่สุดนี้ผู้วิจัยหวังว่างานวิจัยนี้จะเป็นประโยชน์แก่ทุกท่าน และนำไปสู่การคิดค้นและความรู้ใหม่ในอนาคตต่อไป

ราชฤทธา รัตนวรกานต์

มหาวิทยาลัยศิลปากร

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฌ

สารบัญรูปภาพ ..... ญ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของงานวิจัย 2

1.3 สมมติฐานของงานวิจัย 2

1.4 ขอบเขตของการวิจัย 3

1.5 ความจำกัดของการวิจัย 3

1.6 ประโยชน์ที่คาดว่าจะได้รับ 3

บทที่ 2 วรรณกรรมที่เกี่ยวข้อง 4

2.1 ความรู้พื้นฐานและวรรณกรรมที่เกี่ยวข้องกับการระบบเรียนรู้เชิงลึก 4

2.2 Convolutional neural network(CNN) 6

2.2.1. Convolutional layer 6

2.2.2. Pooling layer 7

2.2.3. Fully connect layer 8

2.3 อัลกอริทึมเชิงวิวัฒนาการ (Evolutionary algorithm) 9

2.4 Island-based genetics algorithm (ISLGA) 10

2.5 วรรณกรรมที่เกี่ยวกับการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการในการเรียนรู้เชิงลึก..... 11
บทที่ 3 วิธีดำเนินการวิจัย 22
3.1 รวบรวมข้อมูลและกำหนดขอบเขตของงาน..... 22
3.1.1 กำหนดโปรแกรมที่จะใช้ในการสร้างการเรียนรู้เชิงลึกสำหรับการวิจัย 22
3.1.2 กำหนดฐานข้อมูลของรูปภาพที่ใช้ในการทดสอบและประเมินความสามารถของระบบ 22
3.2 ศึกษาทฤษฎีที่เกี่ยวข้องกับการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการกับการเรียนรู้เชิงลึก..... 23
3.2.1. ออกแบบและทดสอบระบบที่สร้างขึ้นจากแนวคิดใหม่ 23
3.2.2 การออกแบบระบบ 23
3.3 สรุปผลการดำเนินงานวิจัยและจัดทำรายงานวิทยานิพนธ์ 24
บทที่ 4 ผลการดำเนินงาน 25
4.1 การทดลอง..... 25
บทที่ 5 สรุป อภิปรายผล และข้อเสนอแนะ 34
5.1 บทสรุปของงานวิจัย..... 34
5.2 ข้อเสนอแนะ 34
รายการอ้างอิง 2
ประวัติผู้เขียน..... 5

สารบัญตาราง

หน้า

ตารางที่ 1 ตารางแสดงการเปรียบเทียบผลการทำงานกับงานวิจัยที่เกี่ยวข้อง..... 33

สารบัญรูปภาพ

หน้า

ภาพที่ 1 รูปภาพแสดงโครงสร้างของ Googlenet(สืบค้นจาก https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html) 5

ภาพที่ 2 รูปภาพแสดงโครงสร้างของ Alexnet(สืบค้นจาก https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know- About.html) 5

ภาพที่ 3 รูปภาพแสดงการทำงานของ CNN(สืบค้นจาก https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html) 6

ภาพที่ 4 รูปภาพแสดง filter ที่ใช้ในการสร้าง feature maps(สืบค้นจาก https://en.wikipedia.org/wiki/Kernel_(image_processing)) 7

ภาพที่ 5 รูปภาพแสดงการทำงานของ Max pooling(สืบค้นจาก https://embarc.org/embarc_mli/doc/build/html/MLI_kernels/pooling_avg.html) 8

ภาพที่ 6 รูปภาพแสดงการทำงานของ Average pooling(สืบค้นจาก https://embarc.org/embarc_mli/doc/build/html/MLI_kernels/pooling_avg.html) 8

ภาพที่ 7 รูปภาพแสดงหลักการทำงานของอัลกอริทึมเชิงวิวัฒนาการ 9

ภาพที่ 8 รูปภาพแสดงหลักการทำงานของ ISLGA 10

ภาพที่ 9 รูปภาพแสดงค่า Parameter ที่ถูกเข้ารหัส(Assunção, Lourenço, Machado, & Ribeiro, 2019) 14

ภาพที่ 10 รูปแสดงโครงสร้างของระบบที่ดีที่สุดที่ได้จากการทดลอง(Assunção et al., 2019) 15

ภาพที่ 11 รูปภาพแสดงการเข้ารหัสส่วนประกอบของระบบ CNN(Wang, Sun, Xue, & Zhang, 2018) 17

ภาพที่ 12 รูปแสดงอัลกอริทึมการวัดความสามารถของระบบแบบใหม่ที่ถูกเสนอขึ้น(Wang et al., 2018) 18

ภาพที่ 13 รูปภาพแสดงแนวคิดการใช้ building block ในการสร้างการเรียนรู้เชิงลึก(Dutta, Liu, Kurup, & Shah, 2018)..... 19

ภาพที่ 14 รูปภาพแสดงหลักการของการเข้ารหัสแบบ CGP(Suganuma, Shirakawa, & Nagao, 2018)..... 20

ภาพที่ 15 รูปภาพโครงสร้างของการเรียนรู้เชิงลึกที่ใช้ในการทดลอง..... 23

ภาพที่ 16 รูปภาพโครงสร้างของระบบการรียนรู้เชิงลึกที่ใช้ในการทดลอง..... 24

ภาพที่ 17 กระบวนการ ISLGA..... 25

ภาพที่ 18 แผนภาพแสดงกระบวนการทดลอง..... 26

ภาพที่ 19 แผนภูมิกล่องแสดงประสิทธิภาพการทำงานของประชากรในการทดลองที่มีการจำกัดจำนวนประชากรไว้ที่ 4 ต่อกลุ่มในช่วงของประชากร 16 รุ่นแรก..... 27

ภาพที่ 20 แผนภูมิแสดงแนวโน้มของค่าความแม่นยำของประชากรในการทดลองที่มีการจำกัดจำนวนประชากรไว้ที่ 4 ต่อกลุ่ม..... 28

ภาพที่ 21 แผนภูมิกล่องแสดงประสิทธิภาพการทำงานของประชากรในการทดลองที่มีการจำกัดจำนวนประชากรไว้ที่ 8 ต่อกลุ่มของประชากร 16 รุ่นแรก..... 28

ภาพที่ 22 แผนภูมิแสดงแนวโน้มของค่าความแม่นยำของประชากรในการทดลองที่มีการจำกัดจำนวนประชากรไว้ที่ 8 ต่อกลุ่ม..... 29

ภาพที่ 23 แผนภาพแสดงประสิทธิภาพการทำงานของประชากรที่ดีที่สุดที่ได้จากการวิวัฒนาการระบบด้วย ISLGA โดยมีประชากรต่อกลุ่ม = 4..... 30

ภาพที่ 24 แผนภาพแสดงประสิทธิภาพการทำงานของประชากรที่แย่ที่สุดที่ได้จากการวิวัฒนาการระบบด้วย ISLGA โดยมีประชากรต่อกลุ่ม = 4..... 31

ภาพที่ 25 แผนภาพแสดงประสิทธิภาพการทำงานของประชากรที่ดีที่สุดที่ได้จากการวิวัฒนาการระบบด้วย ISLGA โดยมีประชากรต่อกลุ่ม = 8..... 31

ภาพที่ 26 แผนภาพแสดงประสิทธิภาพการทำงานของประชากรที่แย่ที่สุดที่ได้จากการวิวัฒนาการระบบด้วย ISLGA โดยมีประชากรต่อกลุ่ม = 8..... 32

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การเรียนรู้เชิงลึก (Deep learning) เป็นส่วนหนึ่งของการเรียนรู้ของเครื่อง (Machine learning) ที่มีพื้นฐานจากโครงข่ายประสาทเทียม (Artificial neural network) การเรียนรู้เชิงลึกมีโครงสร้างที่หลากหลายและแตกต่างกันจำนวนมาก เช่น Deep neural network (DNNs), Recurrent neural networks (RNNs), Convolutional neural networks (CNNs) และอื่นๆ

ในปัจจุบันการเรียนรู้เชิงลึกเริ่มเป็นที่รู้จักอย่างแพร่หลาย จากความสามารถที่สูงในการประยุกต์ใช้กับงานต่างๆ ในโลกยุคปัจจุบัน โดยเฉพาะ ความสามารถในการรับรู้และแยกแยะเสียงพูดในธรรมชาติ (Natural voice and speech recognition) , การประยุกต์ใช้งานร่วมกับระบบการมองเห็นของคอมพิวเตอร์ (Computer vision) ในช่วงระยะแปดปีที่ผ่านมาการศึกษาเกี่ยวกับการเรียนรู้เชิงลึกมีความก้าวหน้าเพิ่มขึ้นอย่างมาก การศึกษาเหล่านี้ทำให้เกิดการเรียนรู้เชิงลึกที่เป็นที่รู้จักมากมาย เช่น GoogleNet, Resnet, Alexnet อย่างไรก็ตาม ความสามารถของระบบเหล่านี้แลกมาด้วยขนาดของระบบที่มีขนาดใหญ่ทำให้การปรับค่า hyper-parameter ต่างๆ ของระบบทำได้ยากและจะต้องอาศัยความเชี่ยวชาญอย่างสูงในการทำกระบวนการเหล่านี้ ด้วยเหตุนี้อัลกอริทึมเชิงวิวัฒนาการ (Evolutionary Algorithm) จึงถูกนำมาประยุกต์ใช้เพื่อให้การหาค่า hyper-parameter และปรับโครงสร้างของระบบนั้นสะดวกและง่ายขึ้น

อัลกอริทึมเชิงวิวัฒนาการ(evolutionary algorithm) เป็นอัลกอริทึมที่ทำการหาคำตอบของปัญหาต่างๆ โดยอาศัยหลักการทางชีววิทยาในเรื่องของการวิวัฒนาการของสิ่งมีชีวิต หลักการทำงานของอัลกอริทึมโดยสรุปคือ อัลกอริทึม~~นี่จะ~~ทำการสร้างประชากรที่มีรหัสพันธุกรรมเป็นคำตอบของปัญหาขึ้นมากลุ่มหนึ่งในรูปแบบสุ่ม จากนั้นประชากรจะถูกวิวัฒนาการโดยในการวิวัฒนาการแต่ละครั้งประชากรที่มีพันธุกรรมที่อ่อนแอจะถูกคัดออกจากประชากรและถูกแทนที่ด้วยประชากรใหม่ อัลกอริทึมนี้เมื่อนำมาประยุกต์ใช้งานกับปัญหาการหาค่า Hyper-parameter ของการเรียนรู้เชิงลึกจะทำให้ระบบที่ได้มีประสิทธิภาพในการปรับตัวกับข้อมูลที่มีการเปลี่ยนแปลงไปมาได้ดีกว่าวิธีการดั้งเดิมเป็นอย่างมาก อย่างไรก็ตามอัลกอริทึมนี้ก็ยังมีข้อจำกัดบางประการ เช่น ค่าใช้จ่ายในการคำนวณสูง และคำตอบที่ได้จากอัลกอริทึมนี่ก็อาจไม่ใช่คำตอบที่ดีที่สุดเสมอไป

ดังนั้นงานวิจัยนี้จึงมีจุดมุ่งหมายเพื่อทำการทดลองการหาค่า Hyper-parameter ของการเรียนรู้เชิงลึกโดยใช้อัลกอริทึมเชิงวิวัฒนาการ โดยในงานวิจัยนี้จะนำอัลกอริทึมเชิงวิวัฒนาการชนิด Island-based Genetics Algorithm (ISLGA) ซึ่งมีลักษณะเด่นคือประชากรของอัลกอริทึมนี้จะถูกแบ่งเป็นกลุ่ม และมีการแลกเปลี่ยนประชากรระหว่างกลุ่ม ทำให้ประชากรที่ถูกวิวัฒนาการภายใต้อัลกอริทึมนี้สามารถคงความหลากหลายของพันธุกรรมได้ ทำให้โอกาสที่จะได้ Hyper-parameter ที่เหมาะสมที่สุดนั้นสูงขึ้น

1.2 วัตถุประสงค์ของงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อการหาวิธีการใหม่ในการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการในการแก้ปัญหาของการเรียนรู้เชิงลึก โดยมีจุดมุ่งหมายหลักของงานวิจัยคือ

1. เพื่อศึกษาทฤษฎีและความรู้พื้นฐานเกี่ยวกับการเรียนรู้เชิงลึกและปัญหาของการเรียนรู้เชิงลึก

2. เพื่อศึกษาทฤษฎีและความรู้พื้นฐานเกี่ยวกับอัลกอริทึมเชิงวิวัฒนาการเพื่อการประยุกต์ใช้ร่วมกับการเรียนรู้เชิงลึก

3. เพื่อพัฒนาวิธีการใหม่ในการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการเพื่อแก้ปัญหาการหาค่าเหมาะสมที่สุด (Hyper-parameter) ของการเรียนรู้เชิงลึก

1.3 สมมติฐานของงานวิจัย

จากการศึกษาพบว่าปัญหาเกี่ยวกับการเรียนรู้เชิงลึกได้ถูกศึกษามานานแล้ว ในช่วงที่ผ่านมาอัลกอริทึมเชิงวิวัฒนาการเป็นอัลกอริทึมที่นักวิจัยหันมาสนใจจากความสามารถในการปรับตัวและวิวัฒนาการของอัลกอริทึม จากการศึกษางานวิจัยที่ผ่านมาพบว่าอัลกอริทึมเชิงวิวัฒนาการถูกนำมาประยุกต์ใช้ในแก้ปัญหาของการเรียนรู้เชิงลึกรวมถึงเพิ่มประสิทธิภาพของการเรียนรู้เชิงลึกไม่ว่าจะเป็นการปรับปรุงโครงสร้าง หรือการปรับแต่งค่า Hyper-parameter ดังนั้นในงานวิจัยฉบับนี้จะนำอัลกอริทึมเชิงวิวัฒนาการที่มีชื่อว่า Island-based genetic algorithm (ISLGA) มาทดลองใช้ในการหาค่า Hyper-parameter ที่เหมาะสมที่สุดสำหรับการเรียนรู้เชิงลึก จากลักษณะพิเศษของอัลกอริทึมนี้คือการแบ่งประชากรเป็นกลุ่มย่อยหลายกลุ่ม และการบังคับให้มีการแลกเปลี่ยนประชากรระหว่างกลุ่มในระหว่างกระบวนการวิวัฒนาการตามช่วงที่กำหนดทำให้อัลกอริทึมนี่สามารถรักษาความหลายหลายของพันธุกรรมในกลุ่มประชากรได้และนอกจากนี้ยังสามารถรักษาคุณภาพของรหัสพันธุกรรมได้

อีกด้วย ซึ่งทำให้ Hyper-parameter ที่ได้ไม่ลู่เข้าสู่ค่าใดค่าหนึ่งเร็วเกินไปและทำให้การวิวัฒนาการสามารถทำต่อไป ส่งผลให้ค่าของ Hyper-parameter ที่ได้มีโอกาสเป็นค่าที่ดีที่สุด

1.4 ขอบเขตของการวิจัย

ขอบเขตของการวิจัยการหาค่าเหมาะที่สุดของการเรียนรู้เชิงลึกโดยใช้อัลกอริทึมเชิงวิวัฒนาการ มีดังนี้

1. ทำการศึกษาแนวคิดการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการในการเข้ารหัสแทนส่วนประกอบของระบบเพื่อทำการวิวัฒนาการการเรียนรู้เชิงลึก

2. ทำการศึกษาและหาวิธีการใหม่ในการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการกับการเรียนรู้เชิงลึกโดยในแนวคิดใหม่ที่ได้จากการศึกษาวรรณกรรมที่เกี่ยวข้อง

3. ทำการทดสอบและประเมินความสามารถของวิธีการนำเสนอที่สร้างขึ้นจากวิธีการที่ได้ด้วยรูปภาพจากฐานข้อมูล CIFAR-10 โดยเปรียบเทียบประสิทธิภาพกับงานวิจัยอื่นๆ ที่เกี่ยวข้อง

1.5 ความจำกัดของการวิจัย

1.วิธีการใหม่ที่ได้จากการวิจัยมีการใช้ทรัพยากรในการคำนวณสูงเนื่องจากในการออกแบบระบบนั้นโดยไม่คำนึงถึงทรัพยากรการคำนวณที่ต้องใช้เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด

2. วิธีการใหม่ที่ได้เป็นวิธีการที่สามารถประยุกต์ใช้งานได้บนเครื่องคอมพิวเตอร์เท่านั้น

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. วิธีการใหม่ที่นำเสนอทำให้การหาค่า Hyper-parameter ที่เหมาะสมที่สุดทำได้อย่างสะดวกและมีประสิทธิภาพมากขึ้น

2. ระบบที่สร้างขึ้นด้วย Hyper-parameter ที่หาได้จากอัลกอริทึมนี้สามารถทำงานได้อย่างมีประสิทธิภาพ

บทที่ 2
วรรณกรรมที่เกี่ยวข้อง

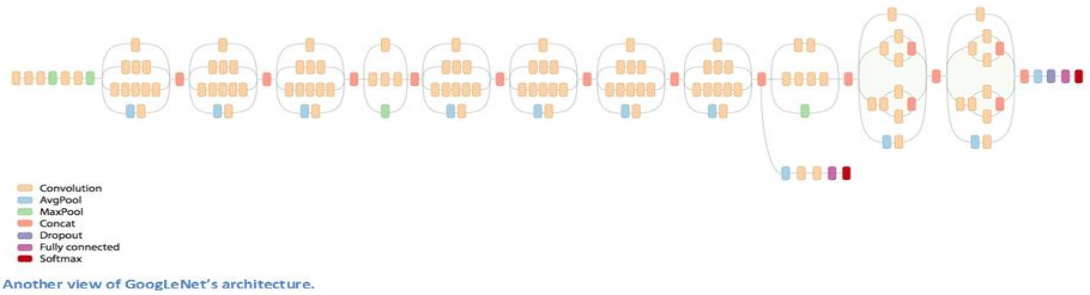
ในส่วนของการทบทวนวรรณกรรมจะทำการนำเสนอความรู้พื้นฐานและงานวิจัยที่เกี่ยวข้องกับการเรียนรู้เชิงลึกโดยเนื้อหาในส่วนนี้ประกอบด้วย ความรู้พื้นฐานและวรรณกรรมที่เกี่ยวข้องกับการระบบเรียนรู้เชิงลึก, อัลกอริทึมเชิงวิวัฒนาการ, Island-based genetics algorithm และวรรณกรรมที่เกี่ยวข้องกับการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการในการหาค่า Hyper parameter ให้กับการเรียนรู้เชิงลึก

2.1 ความรู้พื้นฐานและวรรณกรรมที่เกี่ยวข้องกับการระบบเรียนรู้เชิงลึก

การเรียนรู้เชิงลึกไม่มีนิยามและโครงสร้างที่เป็นที่ตกลงแน่นอนตามที่ Schmidhuber, J. กล่าวไว้ในงานวิจัย Deep Learning in Neural Networks: An Overview ว่า ไม่มีกฎหรือข้อตกลงที่ชัดเจนในการแบ่งแยกประเภทของระบบการเรียนรู้แบบ Shallow network ออกจากการเรียนรู้เชิงลึกแต่นักวิจัยส่วนใหญ่ตกลงกันว่าการเรียนรู้เชิงลึกคือระบบที่ประกอบด้วย Non-linear layer ตั้งแต่2 ชั้นขึ้นไป หรือ 10 ชั้นในกรณีของ Schmidhuber

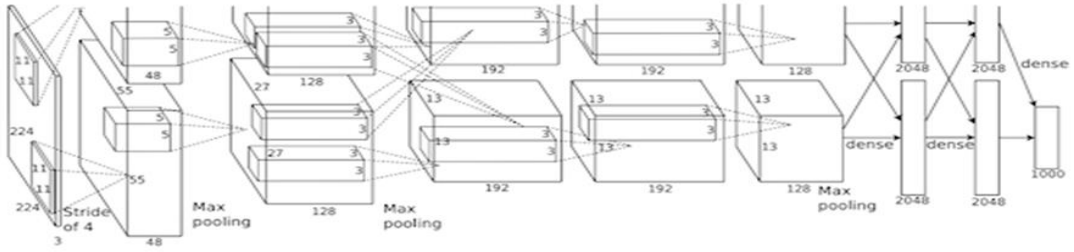
จากการศึกษาพบว่าการเรียนรู้เชิงลึกในยุคเริ่มแรกประกอบด้วย layer เพียงแค่ 3 layer เท่านั้น(Schmidhuber, 2015) ในยุคต่อมาระบบเริ่มมีความซับซ้อนมากขึ้น ระบบเริ่มมีขนาดที่ใหญ่และมีความสามารถและความซับซ้อนมากขึ้น

ตามที่ได้กล่าวไว้ในหัวข้อที่มาและความสำคัญของปัญหาว่าในปัจจุบันการเรียนรู้เชิงลึกหรือ Deep learning เริ่มเป็นที่รู้จักกันอย่างแพร่หลายในช่วงที่ผ่านมาและในปัจจุบันจากความสามารถที่สูงในการประยุกต์และใช้งานที่หลากหลาย เช่น ความสามารถในการรับรู้และแยกแยะเสียงในธรรมชาติ (Natural voice and speech recognition) , การประยุกต์ใช้งานร่วมกับระบบการมองเห็นของคอมพิวเตอร์ (Computer vision) และอื่นๆ อีกมากมาย ทำให้การเรียนรู้เชิงลึกถูกศึกษาค้นคว้าและพัฒนาอย่างต่อเนื่อง การศึกษาเหล่านี้ทำให้เกิดการเรียนรู้เชิงลึกที่เป็นที่รู้จักมากมายหลายระบบ เช่น GoogleNet, Resnet, Alexnet



Another view of GoogLeNet's architecture.

ภาพที่ 1 รูปภาพแสดงโครงสร้างของ Googlenet(สืบค้นจาก https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html)



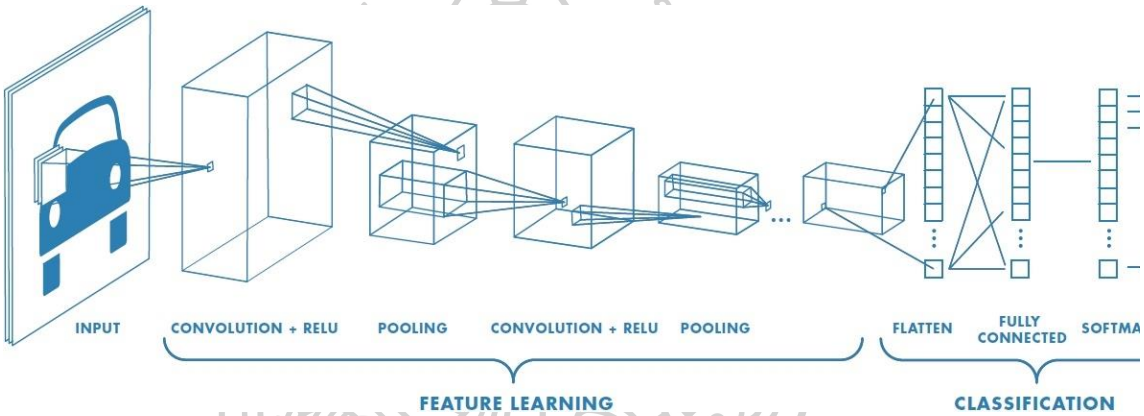
AlexNet architecture (May look weird because there are two different "streams". This is because the training process was so computationally expensive that they had to split the training onto 2 GPUs)

ภาพที่ 2 รูปภาพแสดงโครงสร้างของ Alexnet(สืบค้นจาก https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know- About.html)

จากรูปภาพเป็นตัวอย่างของการเรียนรู้เชิงลึกที่เป็นที่รู้จักกันอยู่ทั่วไป จากรูปแสดงให้เห็นว่าการเรียนรู้เชิงลึกนั้นมีโครงสร้างและมีรูปแบบที่หลากหลาย และมีลักษณะโครงสร้างที่แตกต่างกัน ยกตัวอย่างเช่น จากรูปภาพที่ 1 และ 2 จะเห็นได้ว่า Googlenet และ Alexnet ทั้งสองระบบเป็นระบบมีพื้นฐานมาจากระบบโครงข่ายประสาทเทียมประเภท Convolution neural network (CNN) แต่ Googlenet มีโครงสร้างที่ซับซ้อนและมีส่วนประกอบมากกว่า Alexnet สังเกตได้จาก Dropout layer และ Concat layer ที่เพิ่มขึ้นมาจาก Alexnet จากข้อสังเกตนี้ทำให้เห็นว่าในการปรับปรุงคุณภาพและขีดความสามารถของการเรียนรู้เชิงลึกนั้นสามารถทำได้หลากหลายไม่ว่าจะเป็นการปรับปรุงโครงสร้าง ปรับปรุงคุณสมบัติของแต่ละ layer ภายในระบบ หรือยิ่งไปกว่านั้นเพิ่มส่วนประกอบใหม่ให้กับระบบ อย่างไรก็ตามในงานวิจัยฉบับนี้จะมุ่งเน้นไปที่การปรับปรุง Hyper-parameter ของการเรียนรู้เชิงลึกที่มีต้นแบบมาจาก Alexnet ที่พื้นฐานมาจากระบบโครงข่ายประสาทเทียมประเภท Convolution neural network (CNN)

2.2 Convolutional neural network(CNN)

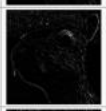
Convolutional neural network หรือ CNN เป็นระบบโครงข่ายประสาทเทียมที่มีลักษณะคล้ายระบบโครงข่ายประสาทเทียมพื้นฐานแต่มีลักษณะเด่นคือ CNN จะทำงานได้ดีกว่ากับ Input ที่ประเภทรูปภาพเนื่องจากความสามารถในการสกัดเอา Feature หรือลักษณะเด่นต่างๆออกมาจากรูปภาพเพื่อใช้ในการเป็น Input ให้แก่ CNN โดยภายในระบบ CNN จะประกอบไปด้วย 3 Layer ได้แก่ Convolutional layer, Pooling layer และ Fully connected layer



ภาพที่ 3 รูปภาพแสดงการทำงานของ CNN(สืบค้นจาก https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html)

2.2.1. Convolutional layer

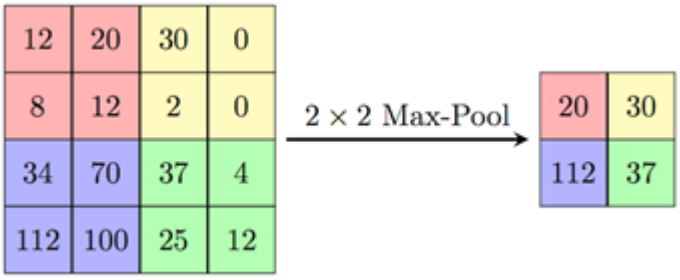
Convolutional layer เป็น Layer ที่ทำหน้าที่ในการสกัด Feature ออกมาจากรูปภาพที่ใช้เป็น input เพื่อนำมาสร้างเป็น Feature map โดยในการสกัด Feature นั้นทำโดยการแบ่งรูปภาพออกเป็นส่วนๆ แต่ละส่วนจะถูกเรียกว่า Cell จากนั้นนำแต่ละ Cell มาผ่านกระบวนการ Convolution กับ Filter เพื่อให้ได้เป็น Feature ออกมาโดย Filter ที่ใช้นั้นขึ้นอยู่กับความต้องการของผู้ใช้ว่าต้อง Feature ใดจากรูปภาพ เช่น ขอบของรูปภาพ เป็นต้น

Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Gaussian blur (approximation)	$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

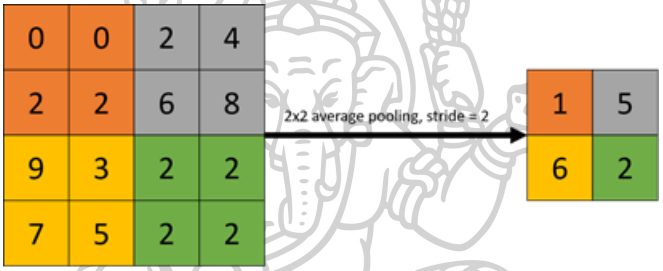
ภาพที่ 4 รูปภาพแสดง filter ที่ใช้ในการสร้าง feature maps(สืบค้นจาก https://en.wikipedia.org/wiki/Kernel_(image_processing))

2.2.2. Pooling layer

Pooling layer เป็น Layer ที่ทำหน้าที่ในการปรับขนาดและปริมาณของข้อมูลตัวอย่าง (Sample) ให้ลดลงก่อนนำส่งเข้าสู่ Layer ถัดไปเพื่อให้สามารถวิเคราะห์และเก็บรายละเอียดของภาพได้อย่างครบถ้วนโดยที่ไม่สูญเสียข้อมูลยิ่งไปกว่านั้นกระบวนการนี้ยังช่วยลดโอกาสเกิดเหตุการณ์ Overfitting ได้อีกด้วย ในการ Pooling นั้นจะมีกระบวนการที่คล้ายกับกระบวนการสร้าง Feature maps คือการแบ่ง Feature map ออกเป็น Cell จากนั้นนำ Cell ไปผ่านกระบวนการ Pooling โดยการทำ Convolution กับ Filter อีกครั้งเหมือนกับ Convolutional layer การ Pooling ที่นิยมกระทำในปัจจุบันมีอยู่สองรูปแบบคือ Max pooling, Average pooling โดยกระบวนการทั้งสองสามารถอธิบายได้ด้วยแผนภาพดังนี้



ภาพที่ 5 รูปภาพแสดงการทำงานของ Max pooling(สืบค้นจาก https://embarc.org/embarc_mli/doc/build/html/MLI_kernels/pooling_avg.html)



ภาพที่ 6 รูปภาพแสดงการทำงานของ Average pooling(สืบค้นจาก https://embarc.org/embarc_mli/doc/build/html/MLI_kernels/pooling_avg.html)

2.2.3. Fully connect layer

Layer นี่ประกอบด้วยระบบ Multilayer perceptron (MLP) ในการประมวลผลข้อมูลที่ได้มาจาก 2 layer ก่อนหน้านี่เพื่อสังเคราะห์และทำการแยกแยะรูปภาพออกเป็นหมวดหมู่

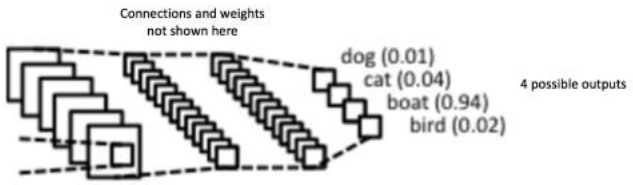
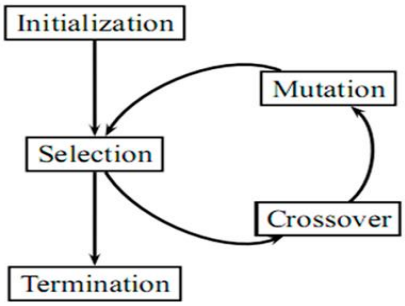


Figure 7 รูปภาพแสดงการทำงานของ Fully connected layer(สืบค้นจาก https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/)

2.3 อัลกอริทึมเชิงวิวัฒนาการ (Evolutionary algorithm)

อัลกอริทึมเชิงวิวัฒนาการเป็นกลุ่มอัลกอริทึมที่ถูกพัฒนาขึ้นมาเพื่อแก้ปัญหาเกี่ยวกับการ Optimization โดยแนวทางในการแก้ปัญหาของอัลกอริทึมนี่มีต้นแบบและแนวคิดมาจากการวิวัฒนาการของสิ่งมีชีวิตกล่าวคือ ประชากรที่เป็นคำตอบของปัญหาสามารถตาย, ผสมพันธุ์, กลายพันธุ์, ฯลฯ ได้เหมือนประชากรของสิ่งมีชีวิตในความเป็นจริง หลักการทำงานของอัลกอริทึมเชิงวิวัฒนาการสามารถได้อธิบายได้ดังแผนภาพด้านล่าง อย่างไรก็ตามอัลกอริทึมที่ต่างกันมีรายละเอียดที่แตกต่างกันแต่ยังมีพื้นฐานของกระบวนการเหมือนเดิม



ภาพที่ 7 รูปภาพแสดงหลักการทำงานของอัลกอริทึมเชิงวิวัฒนาการ

อัลกอริทึมนี้ประกอบด้วยกระบวนการหลัก 3 กระบวนการได้แก่

1. Initailization

กระบวนการนี้เป็นกระบวนการเริ่มต้นของอัลกอริทึมโดยกระบวนการนี้เป็นกระบวนการสร้างประชากรกลุ่มแรกขึ้นมาโดยในรหัสพันธุกรรมของประชากรจะมีตัวแปรหรือค่าคำตอบของปัญหาอยู่ เช่น ในกรณีของงานวิจัยนี้รหัสพันธุกรรมของประชากรจะประกอบไปด้วยค่า Hyper-parameter ในการเรียนรู้เชิงลึกที่ต้องการปรับปรุง

2. Selection

ในกระบวนการนี้ประชากรส่วนหนึ่งจะถูกเลือกให้ทำการผสมพันธุ์และกลายพันธุ์เปรียบเสมือนการผสมพันธุ์ของสิ่งมีชีวิตในธรรมชาติซึ่งทำให้รหัสพันธุกรรมของประชากรที่ถูกเลือกมีการแลกเปลี่ยนและผสมกันเกิดเป็นประชากรที่มีรหัสพันธุกรรมใหม่ที่มีคุณสมบัติแตกต่างไปจากเดิม

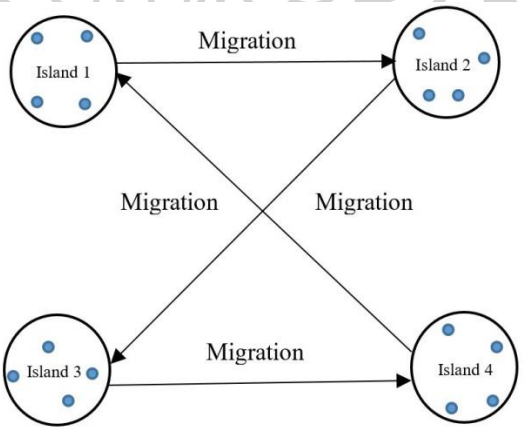
3. Termination

กระบวนการนี้เป็นกระบวนการสุดท้ายในอัลกอริทึมนี้ก่อนที่อัลกอริทึมจะทำกระบวนการในข้อที่ 1 และ 2 ใหม่ กระบวนการนี้เป็นการกำจัดประชากรที่อ่อนแอที่สุดออกไปแล้วทำการแทนที่ด้วยประชากรที่ใหม่ที่ได้จากกระบวนการในข้อที่ 2 ทำให้ประชากรที่จะทำการวิวัฒนาการใหม่ในรุ่นต่อไปมีคุณสมบัติ, ความสามารถที่สูงขึ้น

จากความสามารถในการวิวัฒนาการตัวเองของอัลกอริทึมนี้ทำให้มันได้รับความนิยมเป็นอย่างมากแต่อย่างไรก็ตามอัลกอริทึมก็ยังมีจุดอ่อนบางประการ เช่น คำตอบที่ได้จากอัลกอริทึมนี่อาจจะไม่ใช่คำตอบที่ดีสุดเสมอไป, ค่าใช้จ่ายในการคำนวณที่ค่อนข้างสูง, ฯลฯ

2.4 Island-based genetics algorithm (ISLGA)

Island-based genetics algorithm หรือ ISLGA เป็นอัลกอริทึมหนึ่งในกลุ่มของอัลกอริทึมเชิงวิวัฒนาการ อัลกอริทึมนี้มีจุดเด่นในเรื่องของการจัดกลุ่มของประชากร ตามที่ได้กล่าวมาในหัวข้อที่ 2.3 ว่าอัลกอริทึมเชิงวิวัฒนาการจะทำงานโดยการสร้างประชากรขึ้นมาจำนวนหนึ่งแล้ววิวัฒนาการประชากรเหล่านั้น ISLGA นั้นก็มีหลักการเดียวกันแต่ประชากรของอัลกอริทึมนี่จะถูกจัดให้อยู่เป็นกลุ่มเสมือนว่าแบ่งแยกประชากรเป็นเกาะ (Island) อย่างชัดเจนดังรูป



ภาพที่ 8 รูปภาพแสดงหลักการทำงานของ ISLGA

ในระหว่างกระบวนการวิวัฒนาการจะมีการแลกเปลี่ยนประชากรระหว่างเกาะเพื่อคงความหลายทางพันธุกรรมทั้งในกลุ่มประชากรทั้งหมดและประชากรภายในเกาะ จากลักษณะและ

กระบวนการที่กล่าวมานี้ทำให้อัลกอริทึมมีข้อดีหลายอย่าง เช่น อัลกอริทึมนี้สามารถทำให้คู่กับอัลกอริทึมชนิดอื่นได้, คำตอบที่ได้จากประชากรไม่ลู่เข้าสู่ค่าใดค่าหนึ่งเร็วเกินไป, พันธุกรรมของประชากรถูกควบคุมโดยอัตโนมัติตลอดเวลาจากกลไกการแลกเปลี่ยนประชากรระหว่างเกาะ

2.5 วรรณกรรมที่เกี่ยวกับการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการในการเรียนรู้เชิงลึก

1. Evolutionary convolutional neural networks: An application to handwriting recognition. (Baldominos, A., Saez, Y., & Isasi, P. (2018))

งานวิจัยนี่เป็นการนำอัลกอริทึมเชิงวิวัฒนาการมาประยุกต์ใช้ในการปรับค่า Parameter ต่างๆของระบบโครงข่ายประสาทเทียมประเภท Convolutional neural network โดยในการปรับค่านั้นผู้วิจัยได้นำค่า Parameter ต่างๆ มาใช้เป็นยีนในการวิวัฒนาการระบบ ค่า Parameter ที่ผู้วิจัยได้ให้ความสนใจในแต่ละ Layer ของระบบมีดังนี้

- Convoulutional layer
  จำนวน Convolutional layer
  ขนาดของ Kernel ใน Convolutional layer แต่ละชั้น
  จำนวนของ Kernel ใน Convolutional layer แต่ละชั้น
  Activation function ใน Convolutional layer แต่ละชั้น
  ขนาดของ Pooling layer หลัง Convolutional layer แต่ละชั้น (ถ้ามี)

- Dense (Fully connected) layer
  จำนวนของ Layer
  ลักษณะการเชื่อมต่อของแต่ละ Layer
  จำนวน neuron ในแต่ละ Layer
  Activation function ในแต่ละ Layer
  Weight regularization ในแต่ละ Layer
  Dropout

-

- Parameter ทั่วไป
  Batch size
  Learning rule
  Learning rate

ระบบในแต่ละรุ่นถูกทดสอบด้วยภาพจากฐานข้อมูล MNIST จากการทดลองพบว่าค่าระบบนี่มีค่า Error อยู่ที่ 0.37% โดยระบบดังกล่าวมีคุณสมบัติคือ เป็นระบบที่มี Convolutional layer 3 ชั้นแต่ละชั้นมีขนาด 64x64, 256x256 และ 256x256 ตามลำดับ มี่ขนาดของ Kernel 4x4,2x2 และ 7x7 ตามลำดับ ทุก Layer ใช้ Linear activation function ยกเว้นในชั้นที่ 2 ที่ใช้ Rectified linear unit (ReLU) ระบบมี Pooling layer ขนาด 6x6 ที่ชั้นสุดท้ายของ Convolutional layer เท่านั้น ส่วน Dense layer นั้นเป็นเพียง Multilayer perceptron จำนวน 1 ชั้นที่มี hidden neuron ทั้งหมด 1,024 ตัวแต่ละตัวใช้ Acticvation function ชนิด ReLU มี Dropout rate 50% ระบบทั้งหมดที่กล่าวมานี่ถูก Train โดยใช้อัลกอริทึม Addmax มี Learning rate 0.001 และ Batch size เท่ากับ 100 ซึ่งสิ่งที่น่าสนใจเกี่ยวกับผลการทดลองนี้ คือ ไม่ใช่แค่ระบบเพียงระบบเดียว ระบบอื่นที่มีลักษณะแตกต่างกันที่เป็นประชากรรุ่นเดียวกับระบบนี้ก็มีค่า Error ที่ใกล้เคียงกัน นอกจากค่า Error แล้วสิ่งที่เหมือนกันอีกอย่างคือระบบในรุ่นนี้จะมีลักษณะร่วมกันอยู่สองอย่างคือ ขนาดของ Convolutional layer และ Activation function ประเภท ReLU กล่าวคือ ระบบทุกระบบที่เป็นประชากรของรุ่นนี้จะมี Convolutional layer อย่างน้อย 1 Layer ที่มีขนาด 256x256 และใช้ Activation function เป็น ReLU

2. Design space exploration of Convolutional Neural Networks based on Evolutionary Algorithms (Al-Hyari, A., Areibi, S)

งานวิจัยนี้มีลักษณะคล้ายคลึงกับงานวิจัยก่อนหน้าแต่มีข้อแตกต่างบางประการ ได้แก่

- Model ของระบบที่ใช้ ในงานวิจัยนี้ระบบที่นำมาใช้เป็นพื้นฐานของการวิวัฒนาการคือการเรียนรู้เชิงลึก Alexnet
- โครงสร้างของยีนที่ใช้เป็น Parameter ของ Alexnet

3. Deep Evolutionary Network Structured Representation (Assunc˜ao, F., Lourenc¸o, N., Machado, P., Ribeiro, B.)

งานวิจัยนี้เป็นงานวิจัยเกี่ยวกับการทำ Deep evolutionary structured representation (DENSER) ซึ่งเป็นศาสตร์เกี่ยวกับ การออกแบบระบบโครงข่ายประสาทเทียมอัตโนมัติโดยใช้อัลกอริทึมเชิงวิวัฒนาการ งานวิจัยมีจุดประสงค์หลักเพื่อนำ DENSER มาแก้ไขปัญหาเกี่ยวกับการออกแบบระบบโครงข่ายประสาทเทียมเนื่องจากการออกแบบระบบโครงข่ายประสาทเทียมด้วยมือนั้นเป็นกระบวนการที่ซับซ้อนและวุ่นวายมาก นอกจากต้องมีความรู้พื้นฐานเกี่ยวกับระบบโครงข่ายประสาทเทียมแล้ว การปรับค่า Parameter ต่างๆ และ การปรับปรุงระบบเพื่อเพิ่มประสิทธิภาพนั้นเป็นเหมือนการลองผิดลองถูก ซึ่ง DENSER จะสามารถทำให้สามารถทำกระบวนการเหล่านี้ได้ด้วยตนเองโดยอาศัยความสามารถของอัลกอริทึมเชิงวิวัฒนาการ

สำหรับการทดลองผู้วิจัยได้นำคุณลักษณะและ Parameter ต่างๆมาทำการ encode ให้กลายเป็นยีนสำหรับการวิวัฒนาการโดย Parameter ที่ใช้ถูกแสดงไว้ดังรูป

มหาวิทยาลัยศิลปากร

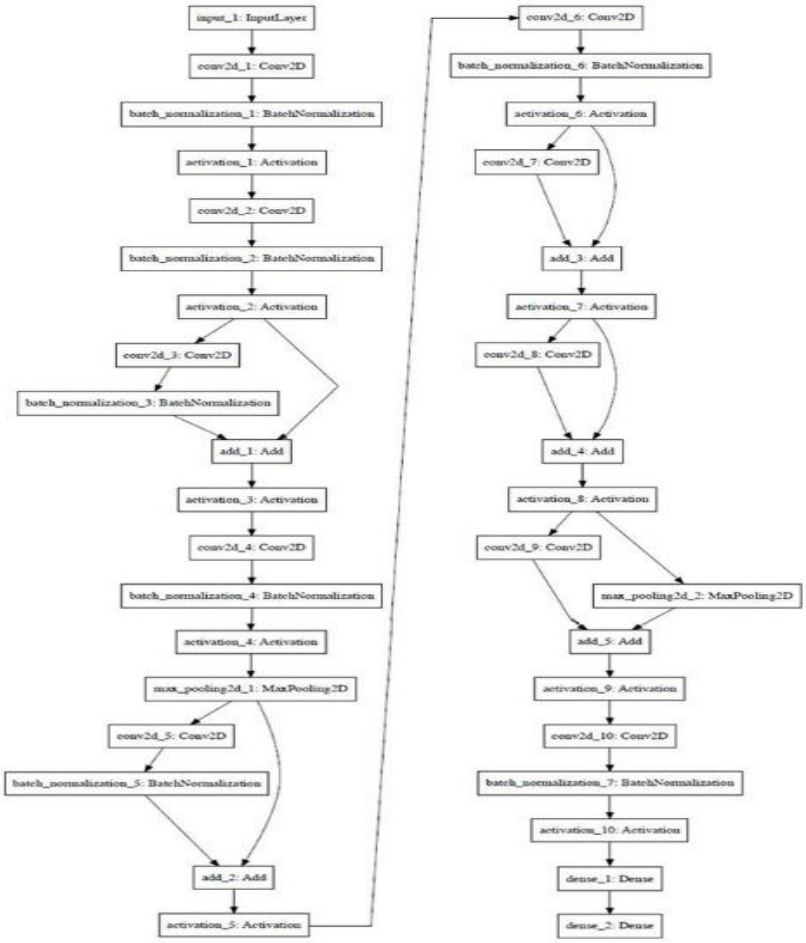
```
<features> ::= <convolution>
             | <pooling>
<convolution> ::= layer:conv [num-filters,int,1,32,256]
                  [filter-shape,int,1,1,5] [stride,int,1,1,3]
                  <padding> <activation> <bias>
                  <batch-normalisation> <merge-input>
<batch-normalisation> ::= batch-normalisation:True
                        | batch-normalisation:False
<merge-input> ::= merge-input:True
                | merge-input:False
<pooling> ::= <pool-type> [kernel-size,int,1,1,5]
              [stride,int,1,1,3] <padding>
<pool-type> ::= layer:pool-avg
              | layer:pool-max
<padding> ::= padding:same
            | padding:valid
<classification> ::= <fully-connected>
<fully-connected> ::= layer:fc <activation>
                      [num-units,int,1,128,2048] <bias>
<activation> ::= act:linear
               | act:relu
               | act:sigmoid
<bias> ::= bias:True
         | bias:False
<softmax> ::= layer:fc act:softmax num-units:10 bias:True
<learning> ::= learning:gradient-descent [lr,float,1,0.0001,0.1]
```

ภาพที่ 9 รูปภาพแสดงค่า Parameter ที่ถูกเข้ารหัส(Assunção, Lourenço, Machado, & Ribeiro, 2019)

ระบบที่ได้จากการทดลองจะถูกทดสอบด้วยการให้ระบบเรียนรู้และแยกแยะรูปภาพจากฐานข้อมูล CIFAR-10 ขนาด 32x32 พิกเซล จำนวน 60,000 รูป

จากการทดสอบพบว่าระบบที่ได้จากการวิวัฒนาการนั้นมีความสามารถที่สูงขึ้นเรื่อยๆ อย่างสังเกตได้ชัด สิ่งที่น่าสนใจคือความสามารถที่เพิ่มมากขึ้นนั้น มาพร้อมกับจำนวน Hidden layer ที่เพิ่มขึ้นโดยระบบในกลุ่มที่มีความสามารถสูงที่สุดนั้นมี Hidden layer เฉลี่ยอยู่ที่ 22.24 ชั้น สำหรับระบบที่มีความสามารถสูงสุดนั้นมีความแม่นยำอยู่ที่ 94.27% โดยมีโครงสร้างของระบบ

ดังนี้



ภาพที่ 10 รูปแสดงโครงสร้างของระบบที่ดีที่สุดที่ได้จากการทดลอง(Assunção et al., 2019)

4. Evolving Deep Convolutional Neural Networks by Variable-length Particle Swarm Optimization for Image Classification (Bin Wang, Yaan Sun, Bing Xue and Mengjie Zhang)

งานวิจัยนี้เป็นงานวิจัยเกี่ยวกับการปรับปรุงระบบโครงข่ายประสาทเทียมชนิด CNN (convolutional neural network) ด้วยอัลกอริทึมเชิงวิวัฒนาการชนิด Particle swarm optimization(PSO) โดยผู้วิจัยมีจุดประสงค์หลักทั้งหมด 3 ประการ ได้แก่

- เพื่อศึกษาวิธีใหม่ในการนำเอาระบบ CNN มาทำการเข้ารหัสและทำการ Optimization ด้วยอัลกอริทึม PSO แบบใหม่
- หาวิธีการใหม่ในการทำลายข้อจำกัดของการเข้ารหัสแบบ Fixed-length เพื่อให้ระบบสามารถปรับขนาดของระบบที่จะสร้างได้
- เสนอวิธีใหม่ในการประเมินความสามารถของระบบโดยการใช้ข้อมูลจาก dataset แค่บางส่วนแทนที่จะใช้ข้อมูล dataset ทั้งหมดเพื่อเพิ่มความเร็วให้กับกระบวนการวิวัฒนาการ

ผู้วิจัยได้เสนออัลกอริทึม PSO แบบใหม่ที่ให้ชื่อว่า IPPSO โดยใช้แนวคิดจากการ ระบบที่อยู่ IP address ที่ใช้ในระบบการสื่อสารข้อมูลมาประยุกต์ใช้งานร่วมกับการเข้ารหัสแบบดั้งเดิม แต่ละชิ้นส่วนของระบบ CNN จะถูกนำมาเข้ารหัสโดยมีขนาดและจำนวน Bit ดังรูป

มหาวิทยาลัยศิลปากร

Layer Type	Parameter	Range	# of Bits	Example Value
Conv	Filter size	[1,8]	3	2(001)
	# of feature maps	[1,128]	7	32(000 1111)
	Stride size	[1,4]	2	2(01)
	Summary		12	001 000 1111 01
Pooling	Kernel size	[1,4]	2	2(01)
	Stride size	[1,4]	2	2(01)
	Type: 1(maximal), 2(average)	[1,2]	1	2(1)
	Place holder	[1,128]	6	32(00 1111)
	Summary		11	01 01 0 00 1111
Fully-connected	# of Neurons	[1,2048]	11	1024(011 11111111)
	Summary		11	011 11111111
Disabled	Place holder	[1,2048]	11	1024(011 11111111)
	Summary		11	011 11111111

ภาพที่ 11 รูปภาพแสดงการเข้ารหัสส่วนประกอบของระบบ CNN(Wang, Sun, Xue, & Zhang, 2018)

จากรูปภาพรหัสดังกล่าวจะถูกนำไปเข้ารหัสอีกครั้งให้เป็นเหมือนรหัส IP address ยกตัวอย่างจากรูปภาพ IP address ที่จะได้จากรหัสดังกล่าวคือ 2.61.18.143.2.61.35.255.27.255 IP address ตามที่ผู้วิจัยศึกษาไว้การกระทำนี้ทำให้ระบบนั้นมีขนาดที่เปลี่ยนแปลงไปมาได้ระหว่างกระบวนการวิวัฒนาการส่งผลให้ความหลากหลายของระบบในประชากรแต่ละรุ่นมีจำนวนมากขึ้น

อีกสิ่งหนึ่งที่ผู้วิจัยพยายามนำเสนอคือวิธีการวัดความสามารถของระบบแบบใหม่โดยวิธีการที่ถูกนำเสนอนั้นสามารถเขียนเป็น Pseudo code ได้ดังนี้

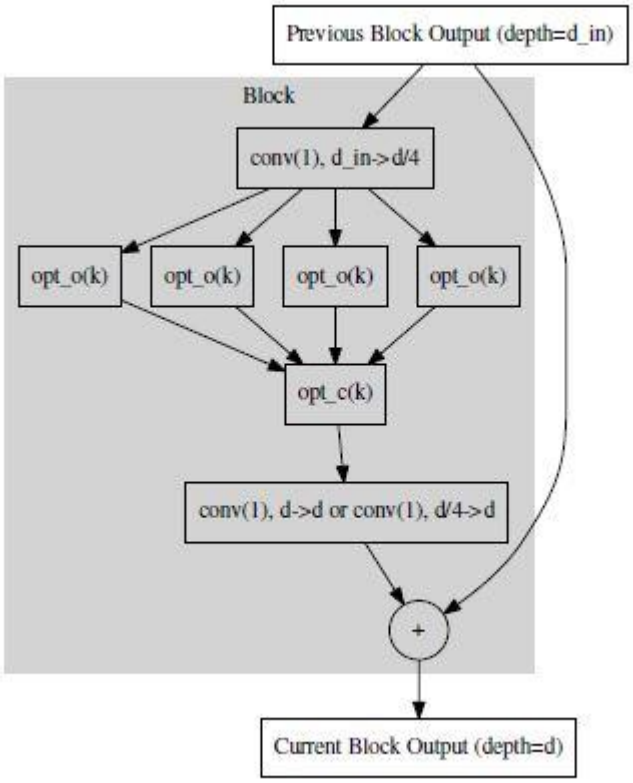
Algorithm 2: Fitness Evaluation

Input: The population P, the training epoch number k, the training set D_{train}, the fitness evaluation dataset $D_{fitness}$, the batch size $batch_size$;
Output: The population with fitness P;
for individual s **in** P **do**
$i \leftarrow 1$;
while $i <= k$ **do**
Train the connection weights of the CNN represented by individual s;
end while
$accy_list \leftarrow$ Batch-evaluate the trained model on the dataset $D_{fitness}$ with the batch size $batch_size$ and store the accuracy for each batch;
$mean \leftarrow$ Calculate the mean value of acc_list
$fitness \leftarrow mean$;
$P \leftarrow$ Update the fitness of the individual ind in the population P;
end for
return P

ภาพที่ 12 รูปแสดงอัลกอริทึมการวัดความสามารถของระบบแบบใหม่ที่ถูกเสนอขึ้น(Wang et al., 2018)

5. Effective Building Block Design for Deep Convolutional Neural Networks using Search.(Dutta, J. K., Liu, J., Kurup, U., & Shah, M. (n.d.))

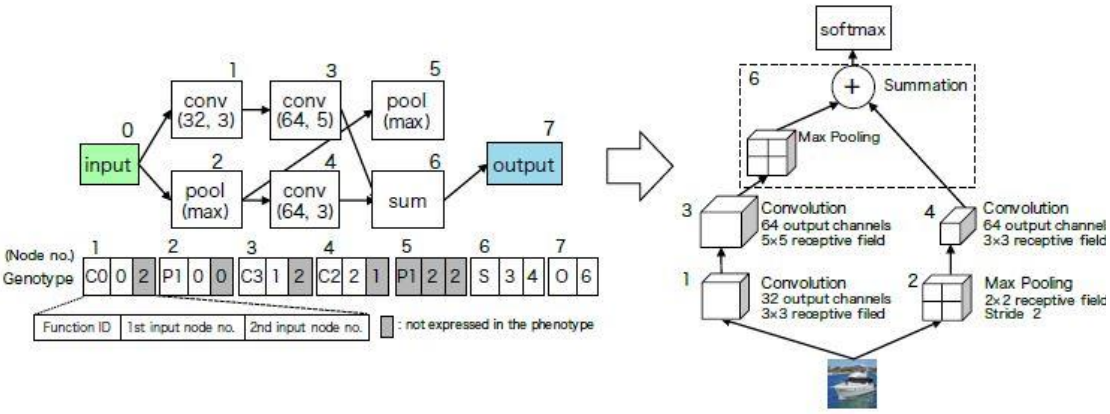
ในการออกแบบการเรียนรู้เชิงลึกอัตโนมัติโดยทั่วไปแล้วจะมีอยู่ 2 วิธีก็คือ การสร้างระบบครั้งเดียวทั้งระบบจากโครงสร้างพื้นฐาน เช่น Convolutional layer, Pooling layer และอื่นๆ เหมือนกับงานวิจัยที่ได้มีการกล่าวถึงก่อนหน้านี่ อีกวิธีหนึ่งคือการมองระบบเป็นเสมือน Block ที่จะถูกใช้เชื่อมต่อกันซ้ำๆเพื่อให้เกิดเป็นการเรียนรู้เชิงลึก วิธีนี้เป็นวิธีการที่มีต้นแบบมาจากการเรียนรู้เชิงลึกแบบ ResNet แต่สิ่งที่งานวิจัยนี่เสนอก็คือการนำวิธีการสร้าง Block แบบ Residue building block ที่มีคุณสมบัติเพิ่มเติมจาก Resnet แบบปกติโดยเพิ่มความสามารถโดยการให้ hyper-parameter ภายในแต่ละ Block สามารถเรียนรู้ด้วยตนเองได้สำหรับการเชื่อมต่อภายใน Block งานวิจัยนี้ใช้อัลกอริทึม Random search แบบทั่วไปในการหาลักษณะการเชื่อมต่อภายในของแต่ละ Block



ภาพที่ 13 รูปภาพแสดงแนวคิดการใช้ building block ในการสร้างการเรียนรู้เชิงลึก(Dutta, Liu, Kurup, & Shah, 2018)

6. A genetic programming approach to designing convolutional neural network architectures. (Suganuma, M., Shirakawa, S., & Nagao, T.)

งานวิจัยนี้มีลักษณะคล้ายกับงานวิจัยงานที่ 4 ตามที่นำเสนอมา กล่าวคืองานวิจัยนี้ได้นำอัลกอริทึมเชิงวิวัฒนาการมาแก้ปัญหาในการสร้างระบบโครงข่ายประสาทเทียมอัตโนมัติที่สามารถปรับขนาดได้ แต่สิ่งที่งานวิจัยนี่ทำแตกต่างจากงานวิจัยก่อนหน้านี้คือวิธีกาสรเข้ารหัส ในงานวิจัยฉบับนี้นำการเข้ารหัสแบบ CGP มาใช้ในการนำชิ้นส่วนต่างๆของระบบมาทำการเข้ารหัสซึ่งการเข้ารหัสด้วนวิธีนี้มีข้อดีคือนอกจากจะทำให้ระบบสามารถปรับค่า Parameter ได้แล้ว ยังทำให้ระบบสามารถแก้ไขการเชื่อมต่อของชิ้นส่วนต่างๆได้ ผู้วิจัยได้แสดงตัวอย่างการทำงานของอัลกอริทึมไว้ดังรูป



ภาพที่ 14 รูปภาพแสดงหลักการของการเข้ารหัสแบบ CGP(Suganuma, Shirakawa, & Nagao, 2018)

อีกสิ่งหนึ่งที่งานวิจัยนี้ทำแตกต่างออกไปคือการใช้อัลกอริทึมเชิงวิวัฒนาการโดยในงานวิจัยนี้จะใช้อัลกอริทึม Evolution strategy (1+λ) ที่ได้รับการปรับปรุง ระบบที่ผ่านการวิวัฒนาการจะถูกทดสอบด้วยรูปภาพจากฐานข้อมูล CIFAR-10 โดยแบ่งเป็นรูปสำหรับการฝึกสอนระบบจำนวน 50,000 รูปและ รูปภาพสำหรับทดสอบระบบจำนวน 10,000 รูป

อย่างไรก็ตามสิ่งที่ผู้เขียนเห็นว่าเป็นปัญหาของงานวิจัยนี้คือ Optimization time ที่นานจากที่ผู้วิจัยรายงานผลไว้พบว่าในการทำ Optimization แต่ละครั้งนั้นต้องใช้เวลามากเฉลี่ยประมาณ 8 – 10 วันและอุปกรณ์ที่ใช้ต้องเป็น GPU ขั้นสูง และยิ่งไปกว่านั้นเมื่อเปรียบเทียบ Optimization กับ Error ที่ได้จากระบบนั้นยังถือว่าสูงพอสมควรเมื่อเปรียบเทียบกันงานวิจัยที่ผ่านมา

7. From Nodes to Networks: Evolving Recurrent Neural Networks (Rawal, A., & Miikkulainen, R.)

งานวิจัยนี้เป็นงานวิจัยที่แตกต่างออกไปจากงานวิจัยก่อนหน้านี้เล็กน้อยงานวิจัยนี้ได้เสนอการการใช้อัลกอริทึมเชิงวิวัฒนาการในการพัฒนาระบบโครงข่ายประสาทเทียมแล้วงานวิจัยยังได้นำเสนอแนวคิดใหม่ในการประเมินคุณภาพของประชากร ซึ่งสิ่งที่น่าสนใจเกี่ยวกับงานวิจัยนี่คือกลยุทธ์ในการตรวจสอบคุณภาพของประชากร ในงานวิจัยนี้ได้นำข้อเสียของการทดสอบระบบด้วยรูปแบบเก่าคือการให้ระบบนั้นวิวัฒนาการแล้วทดสอบระบบด้วยการให้ระบบ Fully trained แต่เนื่องจากวิธีการนี่เป็นวิธีการที่กินเวลาและทรัพยากรค่อนข้างมากจากตัวอย่างที่ผู้วิจัยได้กล่าวไว้ว่า

งานวิจัยก่อนหน้าตนต้องใช้ GPU ถึง 800 อันในการฝึกสอนระบบเพื่อการประเมินประสิทธิภาพแต่ในงานวิจัยฉบับนี้ได้นำเสนอวิธีการใหม่คือการทำนายความสามารถก่อนที่ระบบที่ระบบจะวิวัฒนาการเสร็จโดยใช้ระบบโครงข่ายประสาทเทียมแบบ LSTM จากการทดลองพบว่าวิธีการนี่เป็นวิธีการที่ได้ผลดีกว่าและรวดเร็วกว่าเนื่องจากการทดสอบระบบนั้นไม่ต้องทดสอบระบบทั้งหมดวิธีการที่มีมาก่อนหน้านี่ซึ่งส่งผลให้วิธีการนี่สามารถถูกนำไปประยุกต์ใช้ได้กับการสร้างระบบที่มีขนาดใหญ่ขึ้นได้

8. Dynamic Optimization of Neural Network Structures Using Probabilistic Modeling (Shirakawa, S., Iwata, Y., & Akimoto, Y.)

ในงานวิจัยนี้ผู้วิจัยได้เสนอวิธีการ optimize ระบบโครงข่ายประสาทเทียมทางอ้อมแบบ Dynamic โดยใช้ โมเดลของความน่าจะเป็น โดยในงานวิจัยนี้ผู้วิจัยได้ใช้โมเดลความน่าจะเป็นที่มีการกระจายตัวแบบแบร์นูลี (Bernouli's distribution) เป็นตัวกลางในการปรับค่าน้ำหนักและค่า Parameter ต่างๆในระบบ

จากการศึกษาพบว่าวิธีการที่นำเสนอนี้มีความสามารถที่สูงกว่าทั้งในแง่ของความสามารถของระบบและ Computational time ของระบบนอกจากนั้นระบบที่ถูกสร้างขึ้นมานี่สามารถถูกใช้กับโมเดลความน่าจะเป็นได้หลายรูปแบบนอกจากแบร์นูลี ซึ่งทำให้วิธีการนี้สามารถถูกนำไปประยุกต์ใช้ได้กับระบบได้หลากหลายมากขึ้น

มหาวิทยาลัยศิลปากร

บทที่ 3
วิธีดำเนินการวิจัย

ในส่วนของวิธีในการดำเนินงานวิจัยนี้ จะกล่าวถึงขั้นตอนการดำเนินงานของงานวิจัยฉบับนี้ ซึ่งสามารถแบ่งขั้นตอนการดำเนินงานได้เป็น 4 ขั้นตอน คือ รวบรวมข้อมูลและกำหนดขอบเขตของงาน ศึกษาทฤษฎีที่เกี่ยวข้องกับการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการกับการเรียนรู้เชิงลึก ออกแบบและทดสอบระบบที่สร้างขึ้นจากแนวคิดใหม่ และสรุปผลการดำเนินงานวิจัยและจัดทำรายงานวิทยานิพนธ์ ซึ่งมีรายละเอียดของแต่ละขั้นตอนการดำเนินงานดังนี้

3.1 รวบรวมข้อมูลและกำหนดขอบเขตของงาน

ในขั้นตอนนี้จะเป็นการศึกษา รวบรวมข้อมูลที่เกี่ยวข้องกับระบบงาน ผลงานที่เกี่ยวข้อง แล้วนำข้อมูลที่ได้มาทำการวิเคราะห์ และกำหนดถึงขอบเขตงานวิจัย ซึ่งมีรายละเอียดดังนี้

3.1.1 กำหนดโปรแกรมที่จะใช้ในการสร้างการเรียนรู้เชิงลึกสำหรับการวิจัย

จากการศึกษางานวิจัยอื่นที่เกี่ยวข้องพบว่าในปัจจุบันมีซอฟต์แวร์และภาษาสำหรับใช้ในการเขียนโปรแกรมและพัฒนาการเรียนรู้เชิงลึกจำนวนมากมาย เช่น python, matlab, java และอื่นๆอีกมากมาย ทำให้ในการวิจัยนี้ต้องมีการกำหนดภาษาหรือโปรแกรมที่ใช้ในการสร้างและทดสอบระบบ โดยในงานวิจัยนี้จะทำการออกแบบและสร้างระบบที่ใช้ในการทดลองด้วยโปรแกรมแมทแลบ (MATLAB) เนื่องจากภายในโปรแกรมแมทแลบมีชุดคำสั่งที่สะดวกใช้งานง่ายและมีระบบการฝึกสอนการเรียนรู้เชิงลึกที่มีประสิทธิภาพและอ้างอิงมาจากงานวิจัยชั้นนำทำให้ผลการทดลองที่ได้มีความแม่นยำ นอกจากนี้ยังช่วงในเรื่องการประหยัดเวลาในการพัฒนาระบบและอัลกอริทึม

3.1.2 กำหนดฐานข้อมูลของรูปภาพที่ใช้ในการทดสอบและประเมินความสามารถของระบบ

ในการทดสอบประสิทธิภาพของการเรียนรู้เชิงลึกที่ถูกสร้างขึ้นจาก Hyper-parameter ที่ได้จากอัลกอริทึมที่เลือกใช้ เพื่อให้การทดสอบและประเมินความสามารถของระบบสามารถเชื่อถือได้และมีมาตรฐาน งานวิจัยฉบับนี้จึงเลือกที่จะทำการทดสอบและประเมินความสามารถของการเรียนรู้เชิงลึกดังกล่าวด้วยฐานข้อมูลรูปภาพคละประเภท CIFAR-10 และเปรียบเทียบผลการศึกษากับงานวิจัยอื่นที่เกี่ยวข้อง

3.2 ศึกษาทฤษฎีที่เกี่ยวข้องกับการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการกับการเรียนรู้เชิงลึก

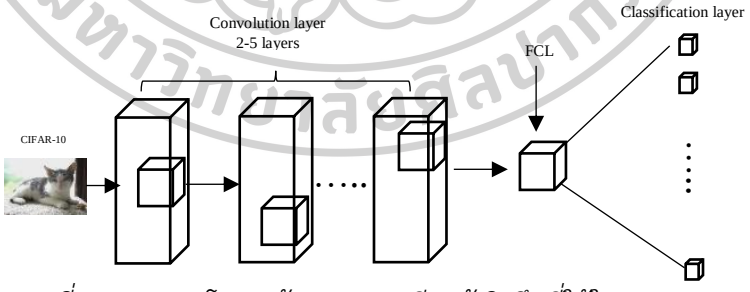
ในขั้นตอนนี้จะเป็นการศึกษาทฤษฎีที่เกี่ยวข้องกับการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการกับการเรียนรู้เชิงลึก เพื่อใช้เป็นแนวทางในการออกแบบและพัฒนาระบบที่จะใช้สำหรับการทดลอง

3.2.1. ออกแบบและทดสอบระบบที่สร้างขึ้นจากแนวคิดใหม่

ขั้นตอนนี้เป็นขั้นตอนในการออกแบบและพัฒนาระบบจากแนวความคิดใหม่ที่ได้จากการวิเคราะห์และสังเคราะห์ความรู้เกี่ยวกับการประยุกต์ใช้อัลกอริทึมเชิงวิวัฒนาการกับการเรียนรู้เชิงลึกที่ได้จากการทบทวนวรรณกรรมและการรวบรวมข้อมูล แล้วนำมาสร้างและพัฒนาเป็นระบบและอัลกอริทึมใหม่ โดยระบบที่สร้างขึ้นจะถูกพัฒนา,ปรับปรุง และทดสอบ ด้วยโปรแกรมแมทแลบ

3.2.2 การออกแบบระบบ

ในงานวิจัยฉบับนี้จะใช้การเรียนรู้เชิงลึกที่มีพื้นฐานมาจากการเรียนรู้เชิงลึกแบบ Alexnet เนื่องจาก Alexnet เป็นการเรียนรู้เชิงลึกที่มีความซับซ้อนไม่มากทำให้สามารถเพิ่มความหลากหลายให้ค่าของ Hyper-parameter โดยที่ไม่เสียค่าใช้จ่ายในคำนวณและฝึกสอนระบบมาก ระบบที่ออกแบบจะมีโครงสร้างเปลี่ยนแปลงไปตามรหัสพันธุกรรมที่ได้จากการใช้อัลกอริทึม ISLGA โดยโครงสร้างของการเรียนรู้เชิงลึกและรหัสพันธุกรรมที่ที่ใช้ในการทดลองสามารถอธิบายได้ดังรูปภาพที่ 14 และ รูปภาพที่ 15



ภาพที่ 15 รูปภาพโครงสร้างของการเรียนรู้เชิงลึกที่ใช้ในการทดลอง

Convolutional layer

<Filter size> : : 1x1 | 3x3 | 5x5 | 7x7 | 9x9

<Number of filter> : : 8 | 16 | 32 | 64 | 128 | 256

<Pooling layer> : : Max pooling | Average pooling

<Active convolutional layer> : : 2 | 3 | 4 | 5

Fully connect layer

<FCL nodes> : : 16 | 32 | 64 | 128 | 256

ภาพที่ 16 รูปภาพโครงสร้างของระบบการรียนรู้เชิงลึกที่ใช้ในการทดลอง

3.3 สรุปผลการดำเนินงานวิจัยและจัดทำรายงานวิทยานิพนธ์

การสรุปผลของการดำเนินงานวิจัยและจัดทำรายงานวิทยานิพนธ์ถือเป็นขั้นตอนสุดท้ายของการวิจัยนี้ โดยจะนำผลลัพธ์ที่ได้จากการทดลองทั้งหมด มาทำการวิเคราะห์และประเมินผลว่าตรงตามจุดประสงค์ที่กำหนดไว้หรือไม่ จากการวิจัยได้ผลการทดลองเป็นที่น่าพอใจ และเป็นไปตามวัตถุประสงค์ที่ตั้งไว้ทุกประการ งานวิจัยนี้ได้ถูกเผยแพร่ในรูปของบทความที่งานประชุมวิชาการระดับนานาชาติ ECTICON 2019 ที่จัดขึ้นในวันที่ 10 – 13 กรกฎาคม 2562 ณ โรงแรมดีวารี จ.ชลบุรี นอกจากนี้บทความที่ได้รับการตอบให้เข้าร่วมจะได้รับการบรรจุเข้าไปในฐานข้อมูล IEEE อีกด้วย

บทที่ 4
ผลการดำเนินงาน

ตามที่ได้กล่าวไว้ในบทก่อนหน้า งานวิจัยนี้เป็นการวิจัยเพื่อศึกษาการหาค่าเหมาะสมที่สุดให้แก่การเรียนรู้เชิงลึกโดยใช้ Island-based genetic algorithm ในหัวข้อนี้จะเป็นการนำเสนอวิธีการทดลองและวิเคราะห์ผลการทดลอง เพื่อวิเคราะห์ถึงประสิทธิภาพของระบบที่ทำการเสนอไว้ในงานวิจัยนี้

4.1 การทดลอง

การทดลองจะถูกทำทั้งหมดแบ่งเป็น 2 ช่วง โดยที่ประชากรทั้งหมดที่ใช้ในกระบวนการวิวัฒนาการจะถูกแบ่งเป็น 4 กลุ่มที่มีจำนวนประชากรเท่ากัน โดยการทดลองจะทำที่จำนวนประกรต่อกลุ่ม 4 และ 8 ประชากรทั้งหมดในการทดลองจะถูกวิวัฒนาการด้วยอัลกอริทึม ISLGA ที่มีหลักการทำงานตาม Pseudo-code ดังรูปภาพที่ 16 และถูกฝึกสอนและทดสอบตามกระบวนการฝึกสอนที่ถูกอธิบายในรูปภาพที่ 17

Algorithm 1 ISLGA algorithm

```
Initialize population size, stopping criteria
FOR each subpopulation P DO
    generate population
    fitness evaluation
    ranking individual by fitness
End DO
generation number := 1
WHILE stopping criteria not met
  FOR each subpopulation P DO
    Crossover
    Mutate random individuals in a subpopulation
    evaluate and select individuals by fitness
    IF[(generation number) mod (frequency) = 0] then
      Send B best individual to a neighboring subpopulation
      Receive B individual from the neighboring subpopulation
    End IF
  Produce new individuals
generation number := generation number +1
End WHILE
```

ภาพที่ 17 กระบวนการ ISLGA

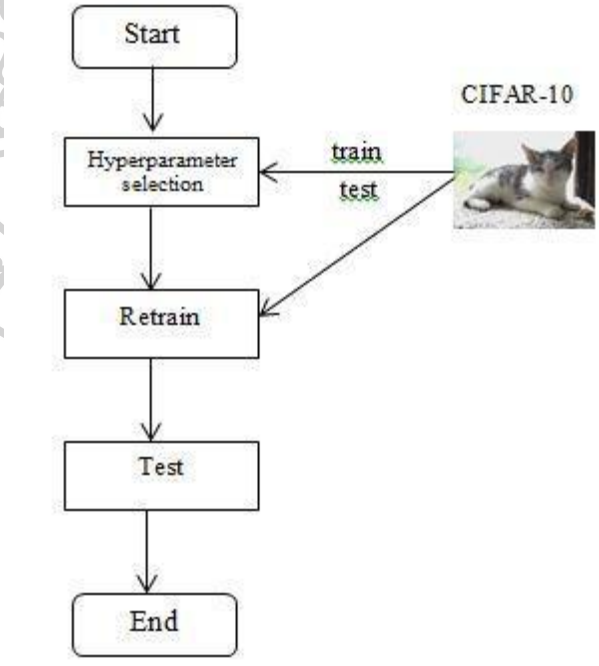
การทำงานของอัลกอริทึม ISLGA ที่ใช้ในการทดลองสามารถอธิบายตามรูปภาพที่ 16 ได้ดังนี้

1. กำหนดประชากรขึ้นมาแบบสุ่มตามจำนวนที่กำหนดโดยประชากรแต่ละตัวมีรหัสพันธุกรรมตามรูปภาพที่ 15

2. ทำการวิวัฒนาการประชากรด้วยอัลกอริทึม ISLGA ตามรูปภาพที่ 16 โดยมีการจัดจำนวนรุ่นสูงสุดไว้ที่ 100 รุ่น ในทุกรุ่นของการวิวัฒนาการประชากรทุกตัวจะถูกทดสอบสมรรถภาพในการทำงานจากการฝึกสอนด้วยรูปภาพจากฐานข้อมูล CIFAR-10 ประชากรที่มีสมรรถภาพต่ำที่สุดในแต่ละกลุ่มจะถูกนำออกและถูกแทนที่ด้วยประชากรใหม่ที่ได้จากการ Crossover ของประชากรที่มีสมรรถภาพสูงสุด 2 ตัวในแต่ละกลุ่ม

3. ในทุกช่วงของการวิวัฒนาการจะมีการกำหนดให้มีการแลกเปลี่ยนประชากรที่มีสมรรถสูงสุดระหว่างกลุ่ม

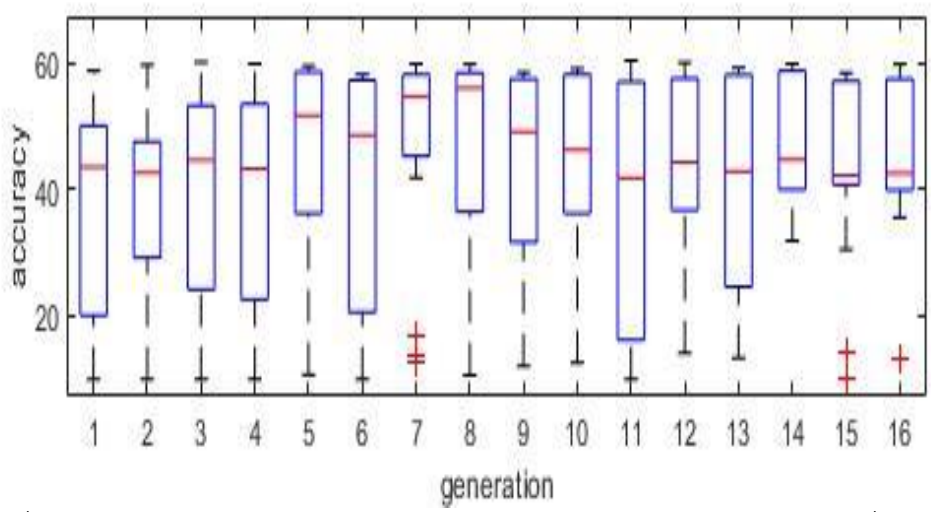
4. หลังจากกระบวนการวิวัฒนาการประชากรที่มีรหัสพันธุกรรมดีที่สุดจะถูกทดสอบอย่างละเอียดอีกครั้งด้วยรูปภาพจากชุดข้อมูลเดิมที่ถูกสลับและเรียบเรียงใหม่



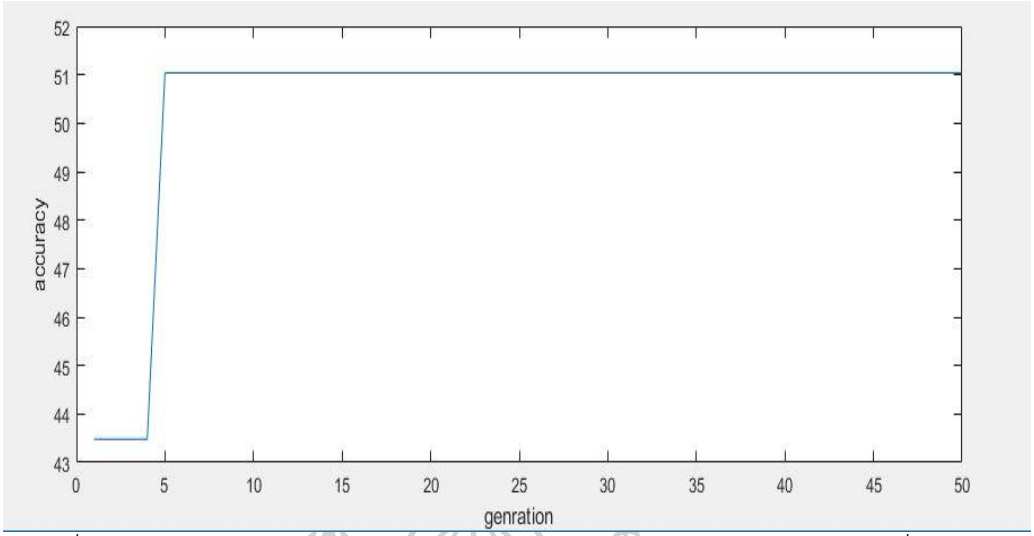
ภาพที่ 18 แผนภาพแสดงกระบวนการทดลอง

จากรูปภาพที่ 17 กระบวนการการฝึกสอนการเรียนรู้เชิงลึกด้วยฐานข้อมูล CIFAR 10 จะเกิดเป็นจำนวนสองครั้งในระหว่างการทดลอง ได้แก่ การฝึกสอนในระหว่างกระบวนการวิวัฒนาการเพื่อทดสอบค่า hyper-parameter เพื่อหาประชากรที่ดีที่สุด และ ประชากรที่จะถูกคัดออก, การฝึกสอนเพื่อทดสอบประสิทธิภาพของตัวแทนที่ครอบครองค่า Hyper-parameter ที่ดีที่สุดหลังจากกระบวนการวิวัฒนาการ โดยกระบวนการฝึกสอนทั้งสองครั้งจะใช้รูปภาพขนาด 32X32 pixel จากฐานข้อมูล CIFAR 10 จำนวน 50,000 รูปในการเป็นชุดฝึกสอน และ 10,000 รูปในชุดทดสอบ ในกระบวนการฝึกสอนระหว่างกระบวนการวิวัฒนาการประชากรแต่ละตัวจะถูกฝึกสอนด้วยอัลกอริทึม Stochastic gradient descent (SGD) ที่มีการกำหนดค่า Momentum = 0 และค่า Batch-size = 50 เป็นจำนวน 50 รอบ(Iteration) และมีการกำหนดจำนวนรุ่นสูงสุดในการทำกระบวนการวิวัฒนาการไว้ที่ 50 สำหรับการทดสอบประสิทธิภาพของตัวแทนที่ดีที่สุด หลังจากกระบวนการวิวัฒนาการเสร็จสิ้นตัวแทนที่ครอบครองค่า Hyper-parameter ที่ดีที่สุดจะถูกทดสอบด้วยกระบวนการฝึกสอนเดิมที่ถูกใช้ในระหว่างกระบวนการวิวัฒนาการแต่เปลี่ยนค่า Momentum ของอัลกอริทึม SGD เป็น 0.9 และเพิ่มจำนวนรอบการฝึกสอนจาก 50 รอบเป็น 100 รอบ

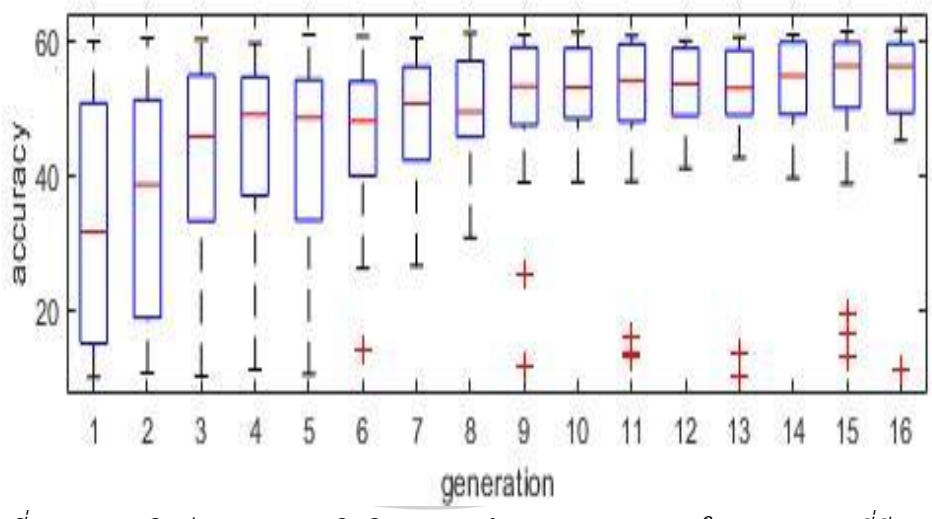
เพื่อการวิเคราะห์ข้อมูลที่มีประสิทธิภาพ ผู้วิจัยได้ทำการเก็บข้อมูลประสิทธิภาพของการเรียนรู้เชิงลึกที่สร้างขึ้นในระหว่างกระบวนการวิวัฒนาการ และ ประสิทธิภาพของระบบการเรียนรู้ที่เป็นตัวแทนที่ดี และแย่ที่สุดในการทดลอง รวมไปถึงการเปรียบเทียบประสิทธิภาพของ hyper-parameter ที่ได้กับผลการทดลองกับงานวิจัยอื่นๆที่เกี่ยวข้อง



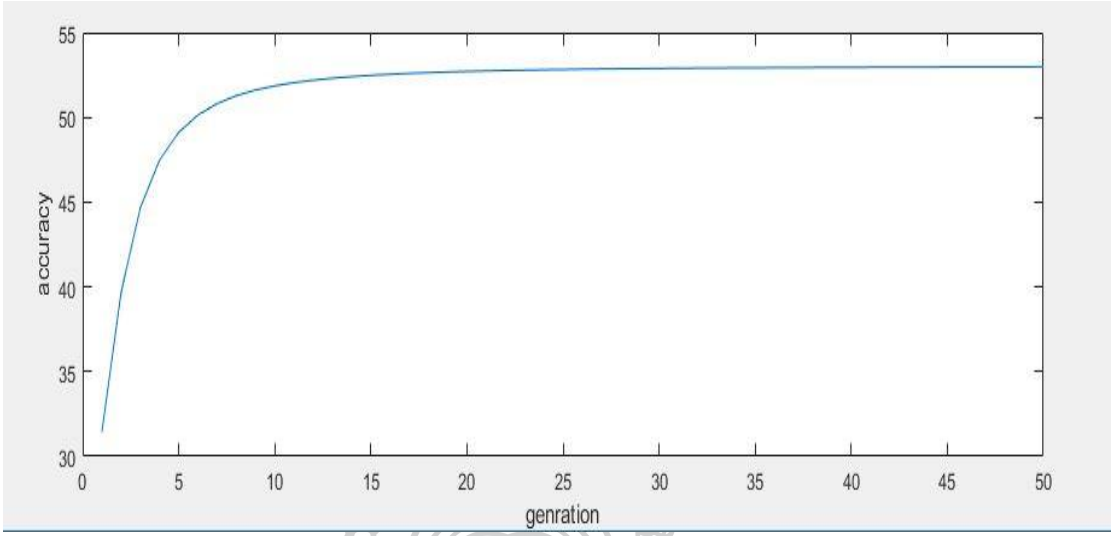
ภาพที่ 19 แผนภูมิกล่องแสดงประสิทธิภาพการทำงานของประชากรในการทดลองที่มีการจำกัดจำนวนประชากรไว้ที่ 4 ต่อกลุ่มในช่วงของประชากร 16 รุ่นแรก



ภาพที่ 20 แผนภูมิแสดงแนวโน้มของค่าความแม่นยำของประชากรในการทดลองที่มีการจำกัดจำนวนประชากรไว้ที่ 4 ต่อกลุ่ม



ภาพที่ 21 แผนภูมิกล่องแสดงประสิทธิภาพการทำงานของประชากรในการทดลองที่มีการจำกัดจำนวนประชากรไว้ที่ 8 ต่อกลุ่มของประชากร 16 รุ่นแรก

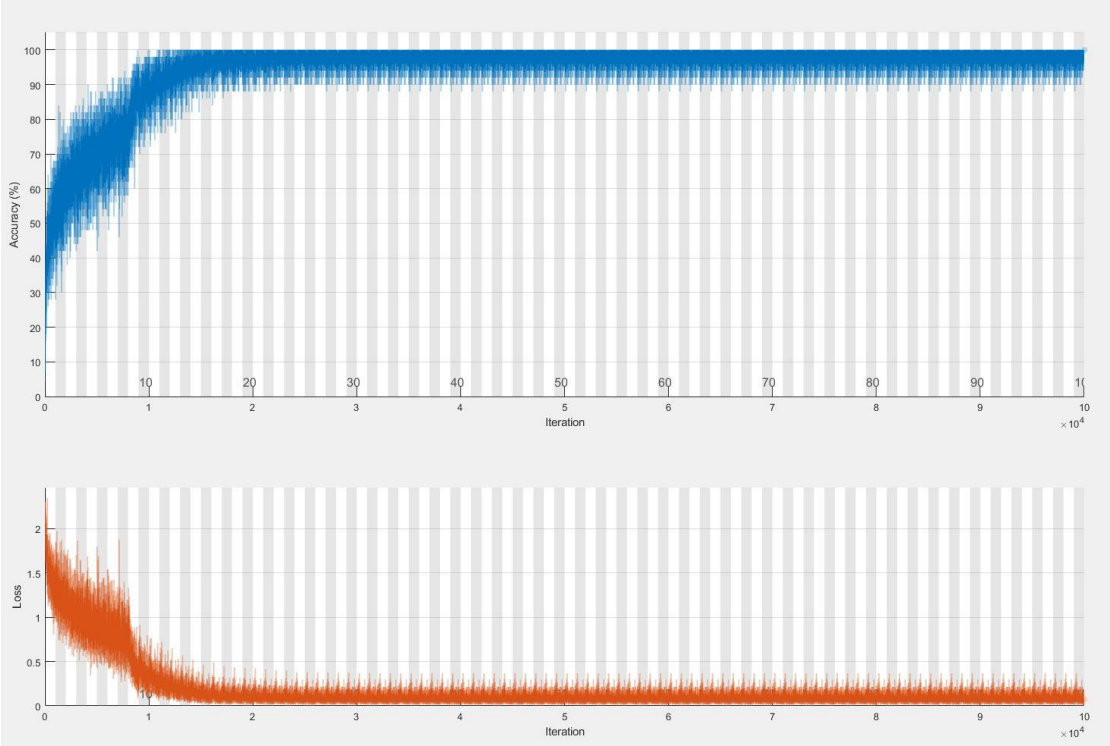


ภาพที่ 22 แผนภูมิแสดงแนวโน้มของค่าความแม่นยำของประชากรในการทดลองที่มีการจำกัดจำนวนประชากรไว้ที่ 8 ต่อกลุ่ม

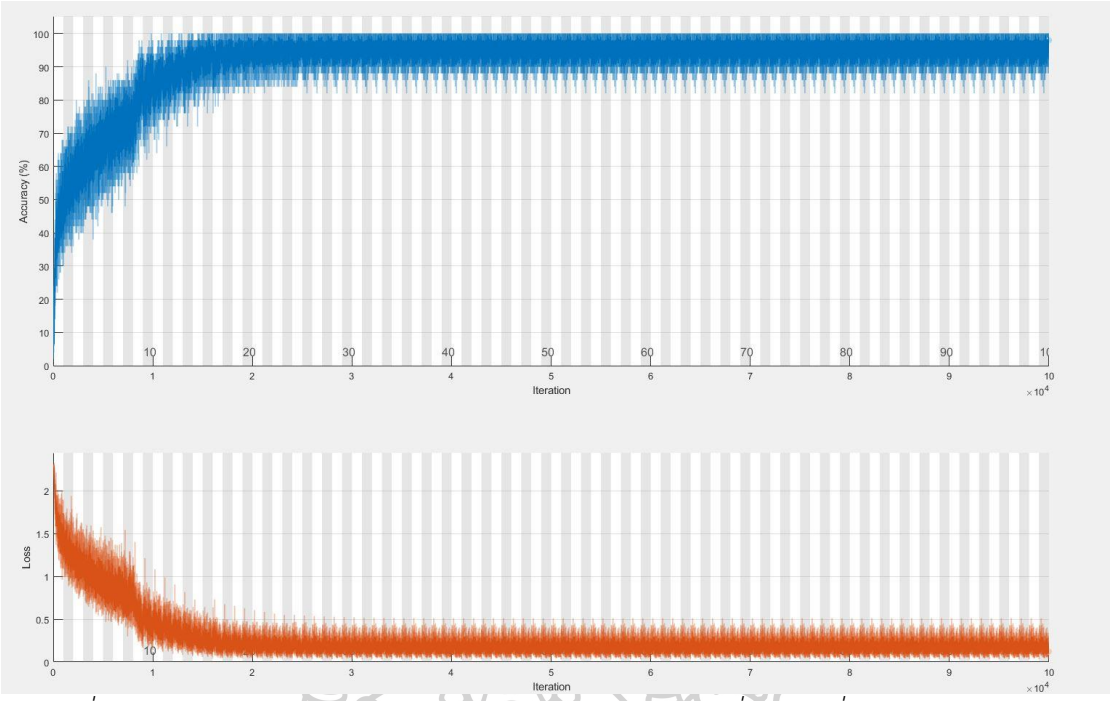
รูปภาพที่ 17 และ 19 เป็นการแสดงความสามารถของประชากรทั้งหมดในการเรียนรู้และแยกแยะรูปภาพจากฐานข้อมูล CIFAR-10 ในประชากร 16 รุ่นแรกจากทั้งหมด 50 รุ่นในรูปของแผนภูมิกล่อง (Box chart) และในรูปภาพที่ 18 และ 20 เป็นการแสดงแนวโน้มของค่าความแม่นยำของระบบของประชากรทั้งหมดตั้งแต่รุ่นแรกถึงรุ่นสุดท้ายโดยกราฟที่ได้สร้างมาจากการนำค่ากลางมาทำการ fixing จากรูปภาพที่ 17 และ 19 จะสังเกตได้ว่า ในช่วงเริ่มต้นของกระบวนการวิวัฒนาการความสามารถในการรู้จำรูปภาพของประชากรจะค่อนข้างกระจัดกระจายโดยมีค่าความแม่นยำอยู่ที่ร้อยละ 20 – 50 ในการทดลองที่จำกัดประชากรไว้ที่ 4 ตัวต่อกลุ่ม และ ร้อยละ 10 – 50 ในการทดลองที่จำกัดประชากรไว้ที่ 8 ตัวต่อกลุ่ม ซึ่งเป็นช่วงที่ค่อนข้างกว้างแสดงให้เห็นถึงความหลากหลายของพันธุกรรมของประชากรเริ่มต้น แต่เมื่อเวลาผ่านไปจะสามารถสังเกตเห็นการเปลี่ยนแปลงของความกระจัดกระจายของค่าความแม่นยำ โดยจะสังเกตเห็นได้ว่ายิ่งจำนวนรุ่นมากขึ้นความกระจัดกระจายของความแม่นยำนั้นลดลงอย่างเห็นได้ชัดโดยแนวโน้มนี้เป็นไปในทิศทางเดียวกันไม่ใช่แค่เพียงประชากร 16 รุ่นที่ยกมาแสดงให้เห็นท่านั้น แต่แนวโน้มนี้เกิดขึ้นจนกระทั่งจบกระบวนการวิวัฒนาการตามรูปภาพที่ 18 และ 20 โดยหลังกระบวนการวิวัฒนาการเสร็จสิ้นค่าความแม่ยำสูงสุดเพิ่มขึ้นเป็นเป็น ร้อยละ 59.52 ในการทดลองที่จำกัดประชากรไว้ที่ 4 ตัวต่อกลุ่ม และ ร้อยละ 61.1 ในการทดลองที่จำกัดประชากรไว้ที่ 8 ตัวต่อกลุ่ม หมายความว่าหาปล่อยให้มีการวิวัฒนาการต่อไปจนมากกว่าจำนวนรุ่นของประชากรที่กำหนด(รุ่นที่ 50) ประชากรที่ได้อาจจะมีความแม่นยำที่สูงขึ้นมากกว่านี้หมายความว่าอัลกอริทึม ISLGA สามารถทำงานได้อย่างมีประสิทธิภาพ นอกจากนั้นจาก

รูปภาพยังแสดงให้เห็นถึงความสามารถในการคงคุณภาพของรหัสพันธุกรรมในกลุ่มประชากรของอัลกอริทึมนี้ หากสังเกตจาก pseudo-code ของอัลกอริทึมนี้ที่ถูกอธิบายด้วยรูปภาพที่ 16 จะสังเกตได้ว่าจากบรรทัดที่ 10 และ 11 ว่าพันธุกรรมของประชากรจะถูกเปลี่ยนชิ้นส่วนในแบบสุ่มทำให้คุณภาพของประชากรมีความแปรปรวนอยู่ตลอดเวลาคล้ายกับกระบวนการการวิวัฒนาการในธรรมชาติ แต่ความสามารถของประชากรก็ไม่เปลี่ยนแปลงไปมากและยังอยู่คงระดับได้ใกล้เคียงกับประชากรกลุ่มใหญ่

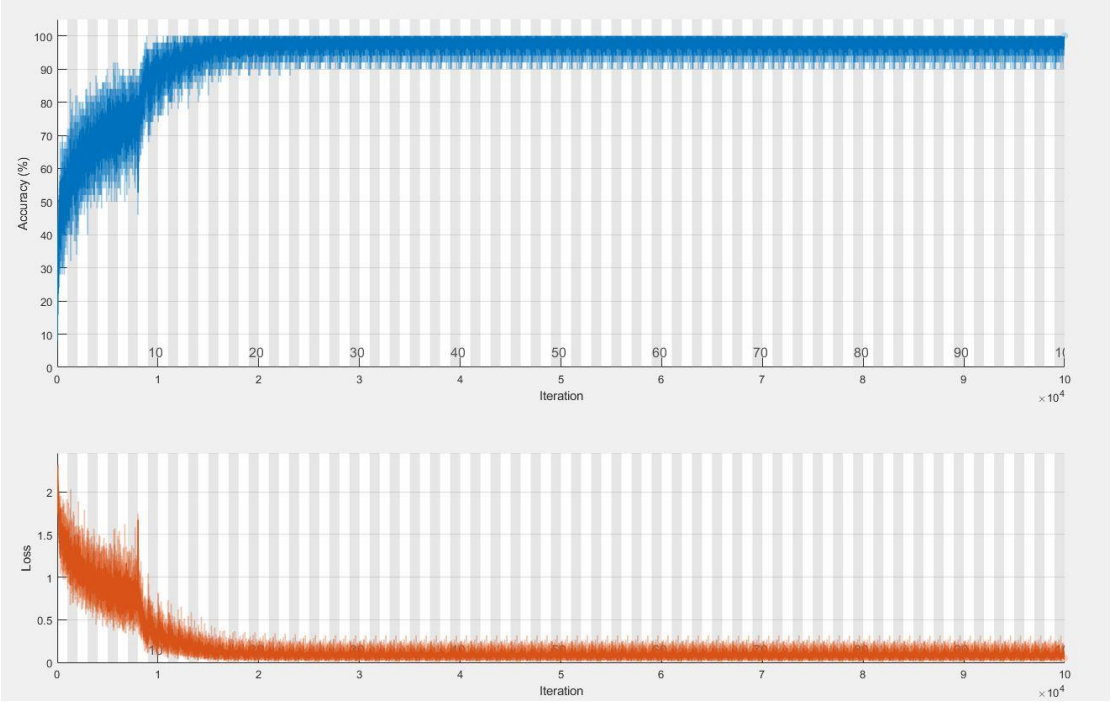
หลังจากการกระวนการวิวัฒนาการเสร็จสิ้นเพื่อเป็นการทดสอบประสิทธิภาพของอัลกอริทึมที่ใช้อย่างละเอียดอีกครั้งผู้วิจัยได้เลือกประชากรที่ดีและแย่ที่สุดในแต่ละการทดลองมาทำการฝึกสอนอีกครั้งโดยใช้ parameter ที่ละเอียดขึ้นตามที่กล่าวไว้ในก่อนหน้านี้ เนื่องจากค่าความแม่นยำที่ได้จากกระบวนการวิวัฒนาการนั้นเป็นค่าที่ได้จากการฝึกสอนแบบหยาบ แผนภาพแสดงความสามารถของประชากรดังกล่าวถูกแสดงไว้ในรูปภาพที่ 21 22 23 และ 24



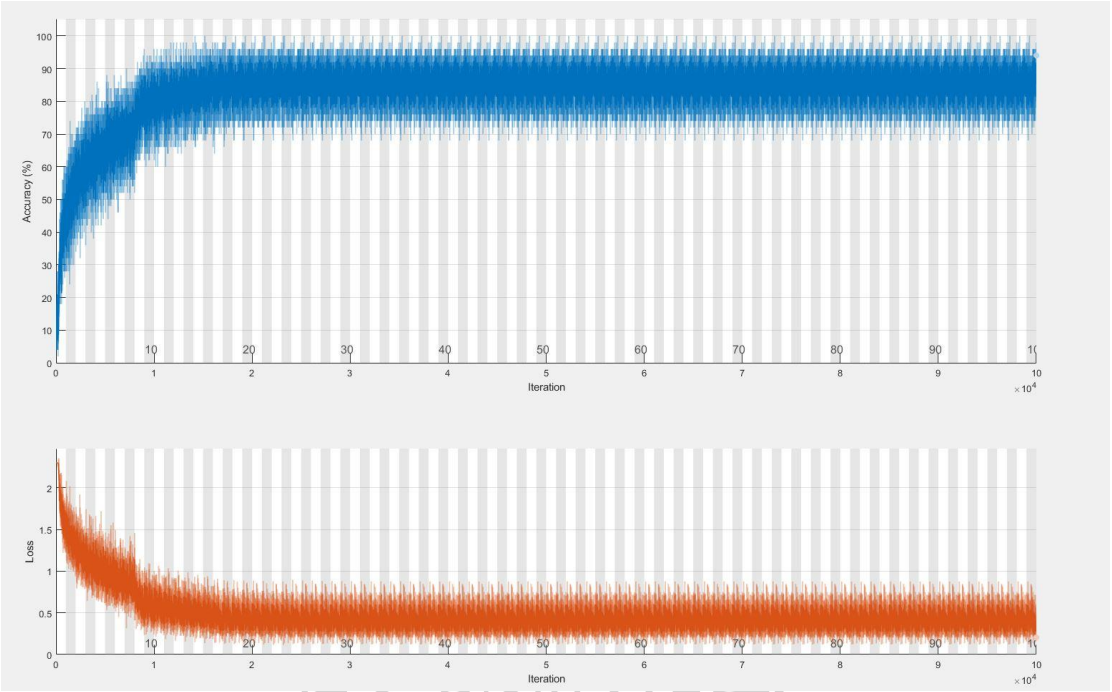
ภาพที่ 23 แผนภาพแสดงประสิทธิภาพการทำงานของประชากรที่ดีที่สุดที่ได้จากการวิวัฒนาการระบบด้วย ISLGA โดยมีประชากรต่อกลุ่ม = 4



ภาพที่ 24 แผนภาพแสดงประสิทธิภาพการทำงานของประชากรที่เย่ทีสุดที่ได้จากการวิวัฒนาการระบบด้วย ISLGA โดยมีประชากรต่อกลุ่ม = 4



ภาพที่ 25 แผนภาพแสดงประสิทธิภาพการทำงานของประชากรที่ดีทีสุดที่ได้จากการวิวัฒนาการระบบด้วย ISLGA โดยมีประชากรต่อกลุ่ม = 8



ภาพที่ 26 แผนภาพแสดงประสิทธิภาพการทำงานของประชากรที่แย่ที่สุดที่ได้จากการวิวัฒนาการระบบด้วย ISLGA โดยมีประชากรต่อกลุ่ม = 8

จากรูปจะสังเกตได้ว่า ประชากรที่ดีสุดมีความสามารถในการรู้จำที่สูงมากเมื่อเปรียบเทียบกับประชากรที่แย่ที่สุดในแต่ละกลุ่มสังเกตได้จากลักษณะของกราฟของค่าความแม่นยำในช่วงรุ่นแรกของประชากรที่แย่ที่สุด(รูปภาพที่ 21 และ 22) ช่วงของความแม่นยำจะมีลักษณะที่กระจัดกระจายมากกว่าประชากรที่ดีที่สุด(รูปภาพที่ 23 และ 24) และนอกจากนี้ช่วงความแม่นยำในระหว่างช่วงของการฝึกสอนที่ประชากรตัวแทนแต่ละตัวสามารถทำได้นั้นยังมีค่าแตกต่างกันอย่างชัดเจน คือ 90 - 100% หรือกล่าวคือระบบสามารถแยกแยะภาพจากชุดทดสอบได้ถูกต้องทั้งหมดในกลุ่มตัวแทนที่ดีที่สุด ในทางกลับกัน 70 – 90% ในกลุ่มประชากรตัวแทนที่แย่ที่สุดซึ่งเป็นความแตกต่างของความสามารถที่เห็นได้อย่างชัดเจน และเมื่อนำค่าความแม่นยำเหล่านี้มาทำการหาค่าเฉลี่ยเพื่อสรุปเป็นค่าความแม่นยำของระบบพบว่าระบบจากการทดลองทั้งสองกลุ่มมีค่าความแม่นยำที่ใกล้เคียงกัน โดยระบบที่ได้จากการทดลองที่มีการจำกัดจำนวนประชากรต่อกลุ่มเท่ากับ 4 มีค่าความแม่นยำ ร้อยละ 73.07 และระบบที่ได้จากการทดลองที่มีการจำกัดจำนวนประชากรต่อกลุ่มเท่ากับ 8 มีค่าความแม่นยำ ร้อยละ 71.34 และเมื่อนำสมรรถนะของระบบที่ได้ไปเปรียบเทียบกับงานวิจัยอื่นๆที่เกี่ยวข้องดังตารางที่ 1

Algorithm	Error rate (%)
CGP-CNN (ResSet) [15]	5.98
CGP-CNN (ConvSet) [15]	6.75
ISLGA-CNN, s = 4	26.93
ISLGA-CNN, s = 8	28.66
SimpleNet-1, s = 4 (PSO) [14]	48.94
SimpleNet-1, s = 10 (PSO) [14]	46.40
SimpleNet-1, s = 16 (PSO) [14]	44.92
SimpleNet-2, s = 4 (PSO) [14]	44.89

ตารางที่ 1 ตารางแสดงการเปรียบเทียบผลการทำงานกับงานวิจัยที่เกี่ยวข้อง

ในตารางที่ 1 เป็นการแสดงการเปรียบเทียบผลการทดลองกับงานวิจัยที่เกี่ยวข้อง จากตารางจะสามารถสังเกตได้ว่าระบบที่สร้างจาก hyper-parameter ของผู้วิจัยนั้น(ISLGA-CNN,s=4 และ ISLGA-CNN,s=8) ถึงแม้จะมีค่าความแม่นยำไม่สูงมากเมื่อเปรียบเทียบกับระบบอื่นๆ ในตาราง แต่ในทางกลับกันเมื่อพิจารณาถึงโครงสร้างแล้วระบบที่ผู้วิจัยสร้างขึ้นนั้นใช้ทรัพยากรในการคำนวณและสร้างระบบน้อยกว่ามาก จากรูปภาพที่ 15 จะเห็นได้ว่าระบบที่สร้างขึ้นนั้นมีความลึกสูงสุดเพียงแค่ 5 ชั้น ซึ่งถือว่าเป็นระบบที่มีความลึกน้อยมากเมื่อเปรียบเทียบกับระบบอื่นที่มีใช้ในปัจจุบันไม่ว่าจะเป็นการใช้ในชีวิตประจำวันหรือ ในการวิจัยต่างๆ

บทที่ 5
สรุป อภิปรายผล และข้อเสนอแนะ

5.1 บทสรุปของงานวิจัย

งานวิจัยศึกษาการหาค่าเหมาะสมที่สุดหรือ Hyper-parameter ให้กับการเรียนรู้เชิงลึกด้วยอัลกอริทึมเชิงวิวัฒนาการ ในงานวิจัยฉบับนี้ผู้วิจัยได้เลือกใช้อัลกอริทึมเชิงวิวัฒนาการที่มีชื่อว่า Island-based genetic algorithm(ISLGA) ซึ่งมีลักษณะเด่นคือการแบ่งประชากรเป็นกลุ่มย่อยหลายกลุ่มก่อนการวิวัฒนาการเพื่อรักษาความหลากหลายทางพันธุกรรมในระหว่างกระบวนการวิวัฒนาการในการทดลองหาค่า Hypermeter โดยผลการทดลองสามารถสรุปได้ดังนี้ อัลกอริทึม ISLGA สามารถทำงานได้มีประสิทธิภาพเมื่อเปรียบเทียบกับอัลกอริทึมอื่น เช่น ใน [14] ใช้อัลกอริทึม Particle swarm optimization (PSO) ในการทำการทดลองลักษณะเดียวกัน แต่อัลกอริทึม ISLGA ที่ผู้วิจัยเลือกนั้นทำงานได้ดีกว่ามากสังเกตได้จากค่า Error rate ที่ต่ำกว่าถึงร้อยละ 15 ถึง 20 อย่างไรก็ตามค่า Hyper-parameter ที่ได้ไม่ใช่เป็นค่าที่ถือว่าเป็นค่าดีที่สุดหรือตายตัวแต่อย่างใด เป็นเพียงแค่ค่า Hyper-parameter ที่ดีที่สุดอัลกอริทึมสามารถหาได้จากกระบวนการวิวัฒนาการเท่านั้น หากทำกระบวนการทั้งหมดใหม่ค่า Hyper-parameter ที่ได้ก็จะเปลี่ยนแปลงไปด้วยเช่นกัน จากการศึกษาเพิ่มเติมพบว่า นอกจากอัลกอริทึมที่แตกต่างกันจะส่งผลต่อผลลัพธ์ที่ได้ยังมีปัจจัยอื่นอีกมากมายที่ส่งผลต่อกระบวนการวิวัฒนาการและค่า Hyper-parameter ที่ได้ เช่น กระบวนการในการฝึกสอนทั้งระหว่างและหลังกระบวนการวิวัฒนาการ, รูปแบบของการทำกระบวนการ Crossover และ Mutation, การจัดการประชากร, ฯลฯ

5.2 ข้อเสนอแนะ

แม้ว่าการทดลองจะประสบผลสำเร็จเป็นที่น่าพอใจแต่การทดลองและระบบที่สร้างได้ในการวิจัยนี้ยังมีข้อจำกัดหลายประการมาก เช่น จำนวนชั้นของ Convolutional layer ในระบบ CNN ที่ใช้มีจำนวนน้อยมากเมื่อเปรียบเทียบกับงานวิจัยที่เกี่ยวข้อง, ทรัพยากรในคำนวณและเวลาที่ค่อนข้างจำกัด, ระบบที่สร้างมาสามารถทำงานได้บนเครื่องคอมพิวเตอร์เท่านั้น ดังนั้นเพื่อการศึกษาที่ดีและถูกต้องมากขึ้นอาจจะต้องมีการเพิ่มทรัพยากรที่ใช้ในการคำนวณและเวลาในการสอนระบบ รวมไปถึงเพิ่มความซับซ้อนของระบบและ Hyper-parameter ให้มากขึ้นเพื่อเป็นการทดสอบความสามารถและดึงศักยภาพของอัลกอริทึมที่เลือกใช้ออกมาให้ได้มากที่สุด

มหาวิทยาลัยศิลปากร

รายการอ้างอิง

[1] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013*(35), 1798-1828.

[2] Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks, 61*, 85-117.

[3] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. doi:10.1038/nature14539

[4] Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Comput., 18*(7), 1527–1554. doi:10.1162/neco.2006.18.7.1527

[5] Whitley, D., Rana, S., & Heckendorn, R. (1998). The Island Model Genetic Algorithm: On Separability, Population Size and Convergence. *Journal of Computing and Information Technology, 7*.

[6] Baldominos, A., Sáez, Y., & Isasi, P. (2018). Evolutionary Convolutional Neural Networks: an Application to Handwriting Recognition. *Neurocomputing, 283*, 38. doi:10.1016/j.neucom.2017.12.049

[7] Alhyari, A., & Areibi, S. (2017). Design space exploration of Convolutional Neural Networks based on Evolutionary Algorithms. *Journal of Computational Vision and Imaging Systems, 3*. doi:10.15353/vsnl.v3i1.162

[8] Assunção, F., Lourenço, N., Machado, P., & Ribeiro, B. (2019). DENSER: deep evolutionary network structured representation. *Genetic Programming and Evolvable Machines, 20*(1), 5-35. doi:10.1007/s10710-018-9339-y

[9] Wang, B., Sun, Y., Xue, B., & Zhang, M. (2018). *Evolving Deep Convolutional Neural Networks by Variable-Length Particle Swarm Optimization for Image Classification*.

[10] Dutta, J., Liu, J., Kurup, U., & Shah, M. (2018). Effective Building Block Design for Deep Convolutional Neural Networks using Search.

[11] Suganuma, M., Shirakawa, S., & Nagao, T. (2017). *A genetic programming approach to designing convolutional neural network architectures*.

[12] Rawal, A., & Miikkulainen, R. (2018). From Nodes to Networks: Evolving Recurrent Neural Networks.

[13] Shirakawa, S., Iwata, Y., & Akimoto, Y. (2018). Dynamic Optimization of Neural Network Structures Using Probabilistic Modeling.

[14] Ribalta Lorenzo, P., Nalepa, J., Kawulok, M., Sanchez, L., & Ranilla, J. (2017). *Particle swarm optimization for hyper-parameter selection in deep neural networks*.

ประวัติผู้เขียน

ชื่อ-สกุล	ราชฤทธา รัตนวรกานต์
วัน เดือน ปี เกิด	11 มีนาคม 2538
สถานที่เกิด	นนทบุรี
วุฒิการศึกษา	วิศวกรรมศาสตร์บัณฑิต วิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยมหิดล วิศวกรรมศาสตร์มหาบัณฑิต วิศวกรรมไฟฟ้าและคอมพิวเตอร์ มหาวิทยาลัยศิลปากร
ที่อยู่ปัจจุบัน	139 หมู่ 1 ตำบลนครปฐม อำเภอเมือง จังหวัดนครปฐม