

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ปรับปรุง XCS สำหรับชุดข้อมูลที่ไม่สมดุลย์

IMPROVE XCS FOR IMBALANCE DATASETS



T117931



ว. 174
ค. 174
2554

เลขหมู่.....
เลขทะเบียน.....**117931**
วัน,เดือน,ปี....**2.2.ค.ค.2554**

b. 12349276
i.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ.2554

KMITL-2011-BN-M-070-101

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

IMPROVE XCS FOR IMABALANCE DATASETS

SORNCHAI UDOMTHANAPONG



A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING IN COMPUTER ENGINEERING
FACULTY OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2011

KMITL-2011-BN-M-070-101

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2011

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Improve XCS for Imbalance Datasets
Student	Mr. Sornchai Udomthanapong
Student ID.	49060705
Degree	Master of Engineering
Program	Computer Engineering
Year	2011
Thesis Advisor	Assoc. Prof. Dr. Boonwat Attachoo

ABSTRACT

Almost of machine learning are developed by assume that their training datasets are well-balanced; however most of real-world datasets are imbalance. So, the performance of machine may be biases toward the majority class and may predict poorly for minority class.

XCS is one of the most powerful learning classifier systems. It combines reinforcement learning and genetic algorithm to create a set of rules representing the extracted knowledge from dataset. The main advantage of this system is to provide condition-action models that represent human-readable patterns. However, not too much public have yet been studied in imbalance dataset.

In this paper, we propose a novel technique to develop XCS deal with imbalance dataset. The proposed technique uses adaptive parameter that are called perception rate for each rule to provide balance learning between major and minor class. In additional, we propose the extend time to applied genetic algorithm to regard the overgeneral problem. The experiment show that the propose technique improve XCS to classify the dataset with height degree of imbalance classes.

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงด้วยดีเนื่องจากกำลังใจและพระคุณอันหาที่สุดมิได้ จากคุณพ่อคุณแม่ ข้าพเจ้าขอสำนึกในพระคุณนี้อย่างเป็นที่สุด

วิทยานิพนธ์นี้จะไม่สำเร็จลุล่วงหากปราศจากแรงผลักดัน และคำแนะนำที่มีประโยชน์ของ รศ.ดร. บุญวัฒน์ อัครฐ อาจารย์ผู้ควบคุมวิทยานิพนธ์ และ รศ.ดร. เอื้อน ปิ่นเงิน อาจารย์ผู้ควบคุมวิทยานิพนธ์ร่วม ข้าพเจ้าขอกราบขอบพระคุณเป็นอย่างสูง

ข้าพเจ้าขอกราบเท้า คุณครูและอาจารย์ทุกท่านตั้งแต่เล็กจนเติบโตใหญ่ ที่ได้มอบวิชาความรู้ให้แก่ข้าพเจ้า รวมทั้งคำสั่งสอนและอบรมให้ข้าพเจ้าเป็นคนดี ข้าพเจ้าขอกราบขอบพระคุณเป็นอย่างสูง

ข้าพเจ้าขอขอบคุณสำหรับกำลังใจ คำแนะนำ และการช่วยเหลือต่าง ซึ่งเป็นประสบการณ์ที่ดีจาก พี่ พรเทพ โรจวสุ พี่ เกรียงศักดิ์ เตมีย์ พี่ สมบัติ ฝอยทอง พี่ ไพฑูร ศรีนิล และห้องวิจัย 804 ทุกท่าน ตลอดห้าปีที่ลาตกระบังข้าพเจ้าจะไม่มีวันลืม

สุดท้ายนี้คุณค่าและประโยชน์อันพึงมีจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอมอบให้กับผู้มีพระคุณทุกท่าน หากวิทยานิพนธ์ฉบับนี้มีข้อผิดพลาดประการใดข้าพเจ้าขอน้อมรับไว้เพียงผู้เดียว

ศรัชย์ อุดมธนาพงศ์

สารบัญ

หน้า	
บทคัดย่อ.....	I
ABSTRACT.....	II
กิตติกรรมประกาศ.....	III
สารบัญ	IV
สารบัญตาราง	VII
สารบัญรูป	VIII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....	2
1.3 สมมุติฐานของการศึกษา.....	2
1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย.....	2
1.5 ขอบเขตของการวิจัย.....	3
1.6 ขั้นตอนของการศึกษา.....	3
1.7 โครงสร้างวิทยานิพนธ์.....	3
บทที่ 2 ระบบจำแนกประเภทแบบเรียนรู้ได้.....	4
2.1 บทนำ.....	4
2.2 ความรู้เบื้องต้นเกี่ยวกับ XCS.....	4
2.2.1 การนำเสนอองค์ความรู้ (Knowledge Representation).....	6
2.2.2 การทำงานของ XCS.....	7
2.3 การเรียนรู้ของ XCS.....	9
2.3.1 ส่วนการเรียนรู้แบบเสริมกำลัง (Reinforcement Learning Component)	9
2.3.2 กระบวนการค้นหาตัวจำแนกประเภท (Discovery Component).....	11
2.3.3 การลบกฎ (Classifier Deletion)	12
2.3.4 Microclassifiers	12
2.3.5 กระบวนการตรวจวัดความซ้ำซ้อนของกฎ (Subsumption Detection)	12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

บทที่ 3 ชุดข้อมูลที่ไม่สมดุล และแนวทางการแก้ไขปัญหาความไม่สมดุลของชุดข้อมูล	14
3.1 ชุดข้อมูลที่ไม่สมดุล	14
3.2 ดัชนีชี้วัดความไม่สมดุล (Degree of Imbalance)	15
3.3 การศึกษาผลกระทบของความไม่สมดุลของชุดข้อมูลต่อการเรียนรู้ของ XCS	15
3.4 ความสัมพันธ์ของการกำหนดค่าตอบแทนสูงสุดค่าตอบแทนต่ำสุดและค่าความผิดพลาดที่ยอมรับได้ กับอัตราส่วนความไม่สมดุลของชุดข้อมูล	18
3.5 ความสัมพันธ์ของการกำหนดค่าอัตราการเรียนรู้กับอัตราส่วนความไม่สมดุล.....	21
3.6 ความสัมพันธ์ของช่วงเวลาในการทำงานของเจเนติกอัลกอริทึม	25
บทที่ 4 การปรับปรุง XCS สำหรับชุดข้อมูลที่ไม่สมดุล	29
4.1 บทนำ	29
4.2 การปรับปรุง XCS ให้สามารถเรียนรู้ชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลมากๆ ได้	29
4.2.1 การคำนวณอัตราส่วนความไม่สมดุลของชุดข้อมูล.....	29
4.2.2 การปรับค่าอัตราการเรียนรู้	31
4.2.3 การขยายช่วงเวลาในการทำงานของเจเนติกอัลกอริทึม	32
4.3 ชุดข้อมูล (Datasets).....	33
4.3.1 ชุดข้อมูลสังเคราะห์ (Simulate Datasets)	33
4.3.2 ชุดข้อมูลจริง (Real Dataset)	34
4.4 การวัดประสิทธิภาพ	40
4.4.1 ค่าเฉลี่ยเรขาคณิต (Geometric mean)	40
4.4.2 k-Fold Cross-Validation technique.....	42
4.5 การทดลอง	42
4.5.1 การทดลองกับชุดข้อมูลสังเคราะห์.....	42
4.5.2 การทดลองกับชุดข้อมูลจริง 2 กลุ่ม.....	44
4.5.3 การทดลองกับชุดข้อมูลจริงที่มีกลุ่มข้อมูลตั้งแต่ 3 กลุ่มขึ้นไป	48
4.6 การเปรียบเทียบผลการทดลอง	50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา^V และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	53
5.1 สรุปผลการวิจัย.....	53
5.2 ข้อเสนอแนะ.....	54
เอกสารอ้างอิง	55
ภาคผนวก ก. งานวิจัยที่ได้รับการตีพิมพ์	58
ประวัติผู้เขียน.....	71



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา ^{VI} และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตาราง	หน้า
3.1 รายละเอียดक्षणของ XCS หลังจากฝึกสอนกับชุดข้อมูลมัลติเพิล็กซ์เซอร์ที่ระดับความไม่สมดุลระดับที่ $i = 7$	17
3.2 เปรียบเทียบค่า prediction payoff และค่า prediction error ที่ได้จากการทดลอง.....	21
4.1 เปรียบเทียบค่า prediction payoff และค่า prediction error ที่ได้จากการทดลอง.....	32
4.2 ตัวอย่างชุดข้อมูลปัญหา Multiplexer 11 บิต	34
4.3 ชุดข้อมูลที่นำมาใช้ในการทดลองเพื่อวัดค่าประสิทธิภาพ	35
4.4 การกระจายตัวของกลุ่มข้อมูลในชุดข้อมูลที่มีตั้งแต่ 3 กลุ่มขึ้นไป.....	38
4.5 รายละเอียดของชุดข้อมูลที่ใช้ที่มี 2 กลุ่มข้อมูล	39
4.6 การประเมินผลการทำนาย (Confusion matrix).....	41
4.7 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูล '1' และส่วนเบี่ยงเบนมาตรฐานของ XCS.....	43
4.8 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูล โดยวิธีการหาค่าเฉลี่ยเรขาคณิต และส่วนเบี่ยงเบนมาตรฐานของ XCS	45
4.9 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูล โดยวิธีการหาค่าเฉลี่ยเรขาคณิต และส่วนเบี่ยงเบนมาตรฐานของ XCS ในการทดลองกับชุดข้อมูล Balance-scale-dist1 และ Glass-dist3 เมื่อผ่านการเรียนรู้ ที่ 500 1000 และ 1500 รอบ	47
4.10 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูล และส่วนเบี่ยงเบนมาตรฐานของ XCS จากการทดลองกับชุดข้อมูลที่มีกลุ่มข้อมูลตั้งแต่ 3 กลุ่มขึ้นไป.....	49
4.11 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูล และส่วนเบี่ยงเบนมาตรฐานของ XCS จากการทดลองกับชุดข้อมูล Glass ที่จำนวนรอบการเรียนรู้ 500 1000 และ 1500	49
4.12 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูลและส่วนเบี่ยงเบนมาตรฐานของ XCS จากการทดลองกับชุดข้อมูลที่มีกลุ่มข้อมูล 2 กลุ่ม.....	50

สารบัญรูป

รูป	หน้า
2.1 การทำงานของ XCS ในรูปแบบของ agent-environment	5
2.2 โครงสร้างของ XCS.....	6
2.3 การทำงาน XCS	7
2.4 กราฟการปรับค่าความถูกต้อง K	10
2.5 การทำงานของเจเนติกอัลกอริทึมใน XCS.....	11
3.1 ตัวอย่างชุดข้อมูลที่ไม่สมดุล	14
3.2 ผลการทดสอบ XCS มาตรฐาน กับชุดข้อมูลปัญหาสมดุลที่ระดับความไม่สมดุลต่างกัน 10 ระดับ	16
3.3 ช่วงเวลาในการเรียนรู้ของ XCS โดยกำหนดค่าอัตราเรียนรู้เป็น 0.2.....	24
3.4 ช่วงเวลาในการเรียนรู้ของระบบ XCS โดยกำหนดค่าอัตราเรียนรู้เป็น 0.1	25
3.5 ช่วงเวลาในการทำงานของเจเนติกอัลกอริทึมกับกฎ 000### : 0 และ 001### : 1 ที่กำหนดช่วงเวลาในการใช้งานเจเนติกเป็น 4	27
3.6 ช่วงเวลาในการทำงานของเจเนติกอัลกอริทึมกับกฎ 000### : 0 และ 001### : 1	27
4.1 ตัวอย่างแสดงการจำลองการทำงานของอุปกรณ์ผสมสัญญาณขนาด 11 บิต เป็นข้อมูลปัญหาสมดุลที่เพิล็กซ์เซอร์ 11 บิต	34
4.2 แบบจำลองตราซังสองแขน.....	35
4.3 กราฟการกระจายตัวของชุดข้อมูล Balance-Scale	36
4.4 กราฟการกระจายตัวของชุดข้อมูล Tao Grid.....	37
4.5 กราฟการกระจายตัวของชุดข้อมูล Tao Grid ในส่วนของพื้นที่ย่อย.....	37
4.6 กราฟการกระจายตัวของชุดข้อมูล Balance-Scale-dist0	48

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ชุดข้อมูลที่ไม่สมดุล คือ ชุดข้อมูลที่มีจำนวนของข้อมูลของกลุ่มใดกลุ่มหนึ่งมากกว่าจำนวนข้อมูลของกลุ่มอื่นหลายๆ เมื่อความแตกต่างของจำนวนข้อมูลระหว่างสองกลุ่มข้อมูลมีค่ามากจะทำให้การเรียนรู้ของเครื่องจักรกลโน้มเอียงไปทางกลุ่มข้อมูลที่มีจำนวนข้อมูลมากจนไม่สามารถจำแนกข้อมูลในกลุ่มที่มีจำนวนน้อยได้ เนื่องจากจักรกลเรียนรู้ได้รับการพัฒนาโดยมีสมมติฐานว่าชุดข้อมูลที่นำมาฝึกสอนนั้นจะต้องมีจำนวนข้อมูลในแต่ละกลุ่มเท่าๆ กัน แต่ในข้อมูลจริงที่ได้มาจากการเก็บข้อมูล โดยส่วนใหญ่แล้วจำนวนข้อมูลในแต่ละกลุ่มจะไม่เท่ากัน ซึ่งเป็นอุปสรรคต่อการเรียนรู้ของระบบจักรกลเรียนรู้ ดังนั้นจึงควรมีการพัฒนาในระบบจักรกลเรียนรู้ให้สามารถใช้งานได้ แม้ว่าชุดข้อมูลที่ใช้ในการฝึกสอนจะเป็นชุดข้อมูลที่มีความไม่สมดุล

ระบบจำแนกประเภทแบบเรียนรู้ได้ เป็นเทคนิคการเรียนรู้ของจักรกลเรียนรู้รูปแบบหนึ่งที่ได้รวมเอาเทคนิคเจเนติกอัลกอริทึม และการเรียนรู้แบบเสริมกำลัง เข้าไว้ด้วยกัน ระบบจำแนกประเภทแบบเรียนรู้ได้จะทำงานโดยอาศัยกฎเป็นพื้นฐาน ซึ่งมีข้อดีคือสามารถบรรยายให้มนุษย์ทำความเข้าใจได้ง่าย ระบบจำแนกประเภทจะประกอบไปด้วยกฎหลายๆ กฎ โดยที่แต่ละกฎจะมีรูปแบบเป็น “if STATE then ACTION” หลักการเรียนรู้ของระบบจำแนกประเภทแบบเรียนรู้ได้เป็นการพัฒนาฐานข้อมูลของกฎที่มีอยู่ให้ได้กฎที่มีประสิทธิภาพ โดยใช้เจเนติกอัลกอริทึมในการสร้างกฎใหม่ๆ ขึ้นมา และใช้วิธีการเรียนรู้ของการเรียนรู้แบบเสริมกำลังซึ่งมีลักษณะการเรียนรู้แบบ Agent-Environment ในการประเมินประสิทธิภาพของกฎที่สร้างขึ้น

Wilson ได้นำเสนอระบบจำแนกประเภทแบบเรียนรู้ได้ที่มีชื่อว่า eXtended Classifier Systems: XCS ในปี 1995 ซึ่งถือว่าเป็นระบบจำแนกประเภทแบบเรียนรู้ได้ที่ดีที่สุดระบบหนึ่งในปัจจุบันมีรายงานผลการวิจัยการนำ XCS ไปประยุกต์ใช้ในงานการจำแนกประเภทข้อมูล และแสดงให้เห็นว่าสามารถแก้ปัญหา well-known problems ได้ดีใน application ต่างๆ เช่น data mining, simulation modeling, robotics และ adaptive control

แต่อย่างไรก็ตาม XCS ก็ยังมีปัญหาการใช้งานร่วมกับชุดข้อมูลบางชนิด เช่น การตรวจจับการบุกรุกทางเครือข่าย (Network Intrusion detection) และการตรวจสอบการหลอกลวงบัตรเครดิต (credit card fraud detection) เป็นต้น ชุดข้อมูลเหล่านี้มีลักษณะเฉพาะร่วมกันอยู่ คือ ความไม่สมดุลกลุ่มข้อมูล สรุปได้ว่าชุดข้อมูลเหล่านี้เป็นชุดข้อมูลที่มีความไม่สมดุล ดังนั้นการพัฒนา XCS ให้

สามารถจำแนกประเภทและเรียนรู้ชุดข้อมูลที่มีความไม่สมดุลได้ จะทำให้สามารถนำ XCS ไปประยุกต์ใช้ในขอบเขตงานวิจัยที่กว้างขวางมากยิ่งขึ้น

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาผลกระทบของปัญหาความไม่สมดุลของชุดข้อมูลต่อการเรียนรู้ของ XCS
2. เพื่อศึกษาหลักการทำงานของ XCS
3. ศึกษาปัจจัยที่ชุดข้อมูลที่มีความไม่สมดุลมีต่อการเรียนรู้ของ XCS และแนวทางในการพัฒนา XCS เพื่อให้สามารถจำแนกกลุ่มข้อมูลที่มีจำนวนน้อยของชุดข้อมูลที่มีความไม่สมดุลได้อย่างมีประสิทธิภาพ
4. เปรียบเทียบประสิทธิภาพของ XCS ที่ได้รับการปรับปรุงกับ XCS มาตรฐาน เพื่อวัดค่าประสิทธิภาพที่เพิ่มขึ้นจากการปรับปรุง

1.3 สมมุติฐานของการศึกษา

เนื่องจากจักรกลเรียนรู้ได้รับการพัฒนา โดยมีสมมุติฐานว่า ชุดข้อมูลที่นำมาฝึกสอนนั้นจะต้องมีจำนวนข้อมูลในแต่ละกลุ่มเท่าๆ กันทำให้การเรียนรู้ของจักรกลเรียนรู้โน้มเอียงเข้าหากลุ่มข้อมูลที่มีจำนวนมาก ทำให้จักรกลเรียนรู้จำแนกกลุ่มข้อมูลที่มีจำนวนน้อยไม่ได้ ดังนั้นการปรับค่าการเรียนรู้ของจักรกลเรียนรู้ให้สมดุลกันระหว่างกลุ่มข้อมูลที่มีข้อมูลจำนวนน้อยและกลุ่มข้อมูลที่มีข้อมูลจำนวนมาก จะทำให้จักรกลเรียนรู้สามารถจำแนกกลุ่มข้อมูลที่ไม่สมดุลได้อย่างมีประสิทธิภาพ

1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย

เพื่อพัฒนา XCS ให้สามารถจำแนกกลุ่มข้อมูลของชุดข้อมูลที่มีความไม่สมดุลได้อย่างมีประสิทธิภาพจำเป็นต้องทำความเข้าใจหลักการและทฤษฎีดังต่อไปนี้

1. หลักการทำงานของ XCS
2. หลักการทดสอบประสิทธิภาพด้วย 10-fold cross-validation technique
3. การหาค่าเฉลี่ยเรขาคณิต

1.5 ขอบเขตของการวิจัย

1. ศึกษาเปรียบเทียบประสิทธิภาพในการเรียนรู้กลุ่มข้อมูลที่มีความไม่สมดุล ด้วยวิธี 10-fold cross-validation technique โดยคำนวณค่าความถูกต้องในการจำแนกกลุ่มข้อมูลด้วยค่าเฉลี่ยเรขาคณิต
2. ชุดข้อมูลที่ใช้ในการสอบมี 3 ชนิด ได้แก่ ชุดข้อมูลสังเคราะห์ ชุดข้อมูลจริงจาก UCI ทั้งที่มีกลุ่มข้อมูล 2 กลุ่ม และมากกว่า 2 กลุ่ม

1.6 ขั้นตอนของการศึกษา

1. ศึกษาผลกระทบของปัญหาความไม่สมดุลของชุดข้อมูลต่อการเรียนรู้ของจักรกลเรียนรู้
2. ศึกษาหลักการการทำงานของ XCS
3. ศึกษาถึงปัจจัยต่างๆ ที่ส่งผลกระทบต่อการเรียนรู้ของ XCS และแนวทางการปรับสมดุลของการเรียนรู้เพื่อให้ XCS สามารถจำแนกกลุ่มข้อมูลของชุดข้อมูลที่มีความไม่สมดุลได้อย่างมีประสิทธิภาพ
4. ทำการทดลองเพื่อทดสอบประสิทธิภาพในการจำแนกกลุ่มข้อมูลของชุดข้อมูลที่มีความไม่สมดุลด้วยระบบจำแนกประเภทแบบเรียนรู้ได้ที่มีการปรับค่าการเรียนรู้ให้สอดคล้องกับอัตราส่วนความไม่สมดุล รวมถึงวิเคราะห์ผลลัพธ์ที่ได้จากการทดลอง
5. สรุปผลการทดลองพร้อมจัดทำบทความวิชาการ และวิทยานิพนธ์

1.7 โครงสร้างวิทยานิพนธ์

- ในวิทยานิพนธ์ฉบับนี้แบ่งเนื้อหาการนำเสนอออกเป็น 5 บทประกอบด้วยเนื้อหาดังต่อไปนี้
- บทที่ 1 กล่าวถึงความเป็นมาและความสำคัญของปัญหา แนวคิดที่นำเสนอเพื่อปรับปรุงโมเดลดั้งเดิม วัตถุประสงค์และขอบเขตของงานวิจัย
- บทที่ 2 กล่าวถึง XCS โดยอธิบายถึงส่วนประกอบและหลักการการทำงานของ XCS
- บทที่ 3 กล่าวถึงความสำคัญของปัญหาความไม่สมดุลของชุดข้อมูล คุณลักษณะของชุดข้อมูลที่ไม่สมดุล และผลกระทบของชุดข้อมูลที่ไม่สมดุลที่มีผลต่อ XCS
- บทที่ 4 กล่าวถึงการปรับปรุง XCS เพื่อให้สามารถเรียนรู้และจำแนกกลุ่มข้อมูลของชุดข้อมูลที่ไม่สมดุลได้ ทำการทดลองและวิเคราะห์ผลการทดลอง
- บทที่ 5 กล่าวถึงการสรุปผลการทดลอง และแนวทางการทำวิจัยต่อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ระบบจำแนกประเภทแบบเรียนรู้ได้

2.1 บทนำ

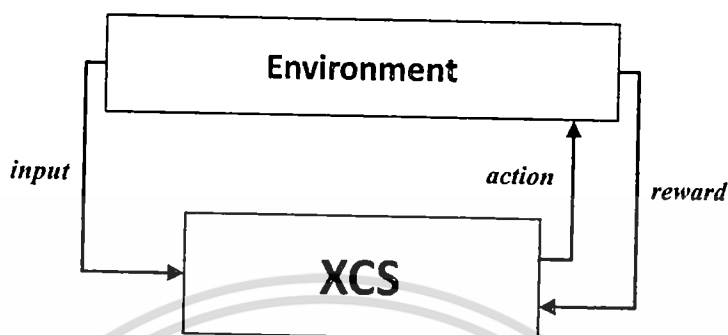
ระบบจำแนกประเภทแบบเรียนรู้ได้ (Learning Classifier System : LCS) [1] คือ ระบบการเรียนรู้ที่อาศัยกฎเป็นพื้นฐาน (rule-based system) โดยผสมผสานเทคนิคในการเรียนรู้ของจักรกลเรียนรู้ 2 ชนิด ได้แก่ การเรียนรู้แบบเสริมกำลัง (Reinforcement learning) และเจเนติกอัลกอริทึม (Genetic algorithm) เข้าไว้ในระบบเดียวกัน ระบบจำแนกประเภทแบบเรียนรู้ได้จะประกอบไปด้วยตัวจำแนกประเภท (classifier) หรือกฎ (rule) เป็นจำนวนมาก โดยแต่ละกฎจะถูกเข้ารหัสไว้อย่างง่ายในรูปแบบ “if CONDITION then ACTION” หลักการเรียนรู้ของระบบจำแนกประเภทแบบเรียนรู้ได้เป็นการพัฒนาฐานข้อมูลของกฎที่มีอยู่ให้ได้กฎที่มีประสิทธิภาพ โดยใช้เจเนติกอัลกอริทึมในการสร้างกฎใหม่ๆ ขึ้นมา และใช้วิธีการเรียนรู้ของการเรียนรู้แบบเสริมกำลังซึ่งมีลักษณะการเรียนรู้แบบ Agent-Environment ในการประเมินประสิทธิภาพของกฎที่สร้างขึ้น

ต้นแบบของระบบจำแนกประเภทแบบเรียนรู้ได้นำเสนอครั้งแรกโดย John Holland [1] ซึ่งเป็นระบบที่มีความซับซ้อนและมีความยุ่งยากในการศึกษาติดตามพฤติกรรมในการเรียนรู้และประสิทธิภาพของระบบ ต่อมา Wilson ได้เสนอ “Zeroth-level Classifier Systems (ZCS)” [2] ซึ่งเป็นระบบที่มีการทำงานคล้ายกับ LCS ต้นแบบของ John Holland แต่สามารถทำความเข้าใจการทำงานได้ง่ายกว่าและยังมีประสิทธิภาพสูงกว่า และต่อมา Wilson ได้พัฒนาระบบใหม่เรียกว่า “eXtended Classifier System – XCS” [3] โดยมีความแตกต่างจาก ZCS ในส่วนของค่าความเหมาะสม (Fitness) ของกฎ มีรายงานวิจัยที่แสดงให้เห็นว่าระบบนี้สามารถแก้ปัญหา well-known problems ได้ดี เช่น data mining, simulation modelling, robotics และ adaptive control [4] ดังนั้นในปัจจุบันนี้ XCS จึงได้รับการยกย่องจากนักวิจัยหลายๆ ท่านว่าเป็นระบบจำแนกประเภทแบบเรียนรู้ได้ที่ดีที่สุดรูปแบบหนึ่ง [5] ดังนั้นในวิทยานิพนธ์นี้จะทำการพัฒนาโดยมี XCS เป็นพื้นฐานในการพัฒนา

2.2 ความรู้เบื้องต้นเกี่ยวกับ XCS

การทำงานของ XCS มีลักษณะการเรียนรู้แบบ Agent-Environment ซึ่งรับอินพุตมาจากสภาพแวดล้อม $s(t) \in S$ และตอบกลับด้วยการกระทำ (Action) $a(t) \in A$ และรับค่าตอบแทน (Reward/Payoff) $R \in \mathcal{R}$ เพื่อเป็นตัววัดค่าผลลัพธ์ของการกระทำที่ส่งออกไป จุดมุ่งหมายของเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

XCS คือ ทำให้ได้มาซึ่งค่าตอบแทนสูงสุด โดยเรียนรู้จากฟังก์ชัน *action-value* ที่ทำการเชื่อมโยงจาก สถานะ – การกระทำ (State-Action) ไปยังค่าตอบแทน $Q(s, a) \rightarrow R$ ซึ่งจะเหมือนกับค่า Q ใน Q -learning ดังรูป 2.1

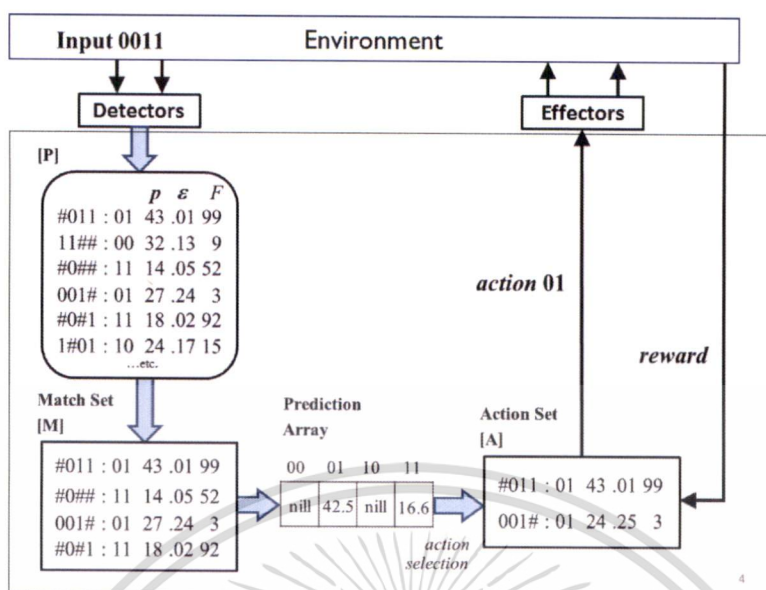


รูป 2.1 การทำงานของ XCS ในรูปแบบของ agent-environment

เนื่องจาก XCS มีการทำงานในรูปของ Agent-Environment ทำให้ XCS สามารถใช้งานกับทั้งการทำงานแบบครั้งเดียว (single-step) คือการทำงานที่ agent จะได้รับค่าตอบแทนจากการทำงานครั้งเดียว เช่น การจำแนกกลุ่มข้อมูล และแบบการทำงานหลายๆ ครั้ง (multi-step) คือการทำงานที่ agent จะได้รับค่าตอบแทนในระยะยาวจากการทำงานหลายๆ ครั้ง เช่น การหาเส้นทางของหุ่นยนต์ ในวิทยานิพนธ์นี้มุ่งเน้นไปที่การใช้งานเกี่ยวกับการจำแนกประเภทของข้อมูล (Classification) ซึ่งเป็นการทำงานแบบครั้งเดียว

โครงสร้างของ XCS จะประกอบด้วย 4 ส่วนด้วยกันได้แก่ เซตประชากร (population set) [P] เซตสอดคล้อง (match set) [M] และ เซตการกระทำ (action set) [A] และ prediction array เซตประชากรเป็นเซตที่เก็บกฎทั้งหมดที่เกิดขึ้นจากการเรียนรู้ของ XCS เซตสอดคล้องเป็นเซตที่เก็บกฎที่สอดคล้องกับอินพุตที่เข้ามาในรอบการทำปัจจุบัน และเซตการกระทำเป็นเซตที่เก็บกฎที่สอดคล้องกับอินพุตและระบุการกระทำหรือกลุ่มข้อมูลที่เลือกเพื่อตอบสนองต่อสภาพแวดล้อม และ prediction array เป็นประมวลผลเพื่อเลือกการกระทำที่จะส่งออกไปยังสภาพแวดล้อม ดังรูป 2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 2.2 โครงสร้างของ XCS

2.2.1 การนำเสนอองค์ความรู้ (Knowledge Representation)

ภายในเขตประชากรจะเก็บกฎที่มีอยู่ทั้งหมดของ XCS กฎเหล่านี้จะแทนองค์ความรู้ที่ได้จากการเรียนรู้ ที่เข้ารหัสไว้ในลักษณะของ เงื่อนไข (Condition C) การกระทำ (Action) และตัวแปรที่ใช้ทำนาย (Prediction parameter) โดยมีรายละเอียดดังนี้

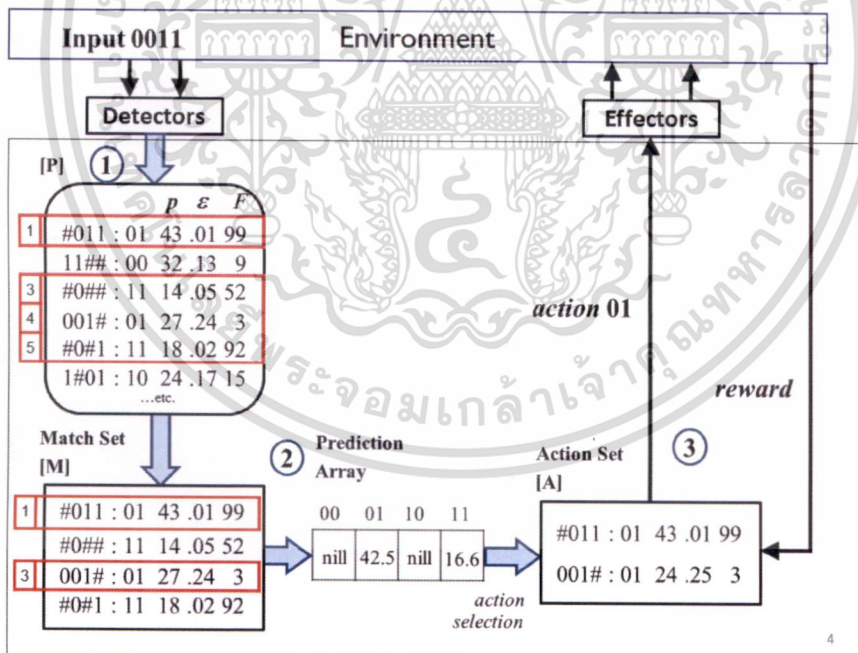
- ส่วนเงื่อนไขสามารถเข้ารหัสได้ 2 แบบ ได้แก่ การเข้ารหัสส่วนเงื่อนไขสำหรับข้อมูลอินพุตที่เป็น binary และการเข้ารหัสส่วนเงื่อนไขสำหรับข้อมูลอินพุตที่เป็น real การเข้ารหัสส่วนเงื่อนไขสำหรับข้อมูลอินพุตที่เป็น binary ซึ่งมียาว L บิต จะเข้ารหัสเป็น $C \in \{0, 1, \#\}^L$ โดยเครื่องหมาย # สามารถแทนได้ทั้ง 0 หรือ 1 (don't care term) ตัวอย่างส่วนของเงื่อนไข เช่น #0#1 จะสามารถสอดคล้องอินพุตที่เข้ามาได้ 4 รูปแบบ ได้แก่ 0001 0011 1001 และ 1011 การเข้ารหัสส่วนเงื่อนไขสำหรับข้อมูลอินพุตที่เป็น real ซึ่งมีจำนวนคุณลักษณะ L ค่า จะเข้ารหัสเป็น $C \in [\text{lower}, \text{upper}]^L$ โดยสามารถสอดคล้องกับข้อมูลที่มีค่าตั้งแต่ lower ไปจนถึง upper
- ส่วนการกระทำ a จะเข้ารหัสเป็น binary string หรือจำนวนเต็มโดยมีจำนวนเท่ากับจำนวนการกระทำทั้งหมดที่เป็นไปได้หรือจำนวนของกลุ่มข้อมูลที่ระบุไว้ (Number of Classes labeled) ในแต่ละกฎจะมีค่าการกระทำได้เพียง 1 ค่าจากค่าการกระทำที่เป็นไปได้ทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ในแต่ละกฎจะมีตัวแปรที่ใช้ทำนาย 3 ตัวมีดังต่อไปนี้
 - ค่าประมาณผลตอบแทน (Prediction Reward) p ซึ่งเป็นค่าประมาณของค่าตอบแทนที่จะได้รับถ้ากฎตรงกับสภาพแวดล้อมที่ได้รับในขณะนั้นและถูกนำไปใช้งาน
 - ค่าประมาณความผิดพลาด ε (Error) ซึ่งเป็นค่าที่ใช้ประมาณของความผิดพลาดในการทำนายค่าตอบแทน กับค่าตอบแทนที่ได้รับจริง
 - ค่าฟิตเนส (Fitness) F ซึ่งเป็นค่าวัดความแข็งแกร่งของกฎนั้น โดยค่าฟิตเนสสามารถคำนวณโดยอาศัยค่า ε

2.2.2 การทำงานของ XCS

การทำงานของ XCS จะเริ่มหลังจากได้รับอินพุตจากสภาพแวดล้อม โดยแบ่งเป็น 3 ขั้นตอนหลักๆ ได้แก่ 1) การค้นหากฎที่สอดคล้อง 2) การเลือกการกระทำ 3) และการสร้างเซตการกระทำ ดังรูป 2.3



รูป 2.3 การทำงาน XCS

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) ขั้นตอนการค้นหากฎที่สอดคล้อง เมื่อ XCS ได้รับอินพุตมาจากสภาพแวดล้อม XCS จะทำการค้นหากฎจากเซตประชากร [P] โดยจะเลือกกฎที่มีส่วนเงื่อนไขที่สอดคล้องกับอินพุตที่เข้ามาตามหัวข้อที่ 2.2.1 และจะนำกฎที่สอดคล้องมารวมกันไว้ในเซตสอดคล้อง [M] จากรูปที่ 2.3 เมื่อ XCS ได้รับอินพุต 0011 XCS จะทำเปรียบเทียบหากฎที่สอดคล้อง โดยจะมีกฎที่สอดคล้อง 4 กฎ ได้แก่ กฎที่ 1 3 4 และ 5 จากเซตประชากร [P]

2) ขั้นตอนการเลือกการกระทำ หลังจากได้เซตสอดคล้อง XCS จะสร้าง prediction array เพื่อใช้ในการเลือกกฎที่จะส่งออกไปยังสภาพแวดล้อม การสร้าง prediction array สร้างจากค่าเฉลี่ยของค่า prediction payoff p โดยมีค่า fitness F เป็นค่าถ่วงน้ำหนัก [3] ดังสมการต่อไปนี้

$$P(A) = \frac{\sum_{a=A \wedge c \in [M]} p \times F}{\sum_{a=A \wedge c \in [M]} F} \quad (2.1)$$

จากรูป 2.3 พบว่าในเซตสอดคล้องมีกฎที่มีการกระทำเป็น 01 ได้แก่ กฎที่ 1 และ 3 จากเซตการกระทำค่า prediction array ของการกระทำ 01 จะมีค่าเป็น $(43 \times 99 + 27 \times 3) / (99 + 3) = 42.5$ และกฎที่มีการกระทำเป็น 11 ได้แก่ กฎที่ 2 และ 4 จากเซตการกระทำ ค่า prediction array ของการกระทำ 11 จะมีค่าเป็น $(14 \times 52 + 18 \times 92) / (52 + 92) = 16.6$

จากนั้น XCS จะเลือกกฎเพื่อส่งออกไปยังสภาพแวดล้อม วิธีการเลือกกฎของ XCS มี 2 วิธี คือ การสำรวจ (explore) และการใช้ประโยชน์ (exploit) การเลือกการกระทำโดยการสำรวจใช้ในเวลาที่ต้องการฝึกสอน XCS โดยการเลือกการกระทำด้วยวิธีสำรวจนี้จะสุ่มเลือกการกระทำอยู่ในเซตสอดคล้อง การเลือกการกระทำโดยการใช้ประโยชน์ จะทำการเลือกการกระทำที่มีค่า prediction ใน prediction array สูงที่สุด

3) ขั้นตอนการสร้างเซตการกระทำ หลังจากเลือกการกระทำที่จะส่งออกไปยังสภาพแวดล้อมแล้ว XCS จะนำกฎที่มีการกระทำตรงกับการกระทำที่เลือกมาไว้ในเซตการกระทำ เพื่อรอรับการปรับค่า พารามิเตอร์ prediction payoff p , prediction error ϵ และค่า fitness F จากรูป 2.3 สมมุติว่า XCS เลือก การกระทำ 01 เพื่อส่งออกไปยังสภาพแวดล้อม ก็จะนำกฎที่ 1 และ 3 จากเซตสอดคล้องมาไว้ในเซตการกระทำ

หลังจากผ่านขั้นตอนทั้งหมดแล้ว XCS จะส่งค่าการกระทำที่เลือกไปยังสภาพแวดล้อม และรอรับค่าตอบแทนกลับมาจากสภาพแวดล้อม เพื่อปรับค่าพารามิเตอร์ พารามิเตอร์ prediction payoff p , prediction error ϵ และค่า fitness F ซึ่งนับได้ว่าจบการทำงานในแต่ละรอบของ XCS

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 การเรียนรู้ของ XCS

การเรียนรู้ของ XCS เป็นการเรียนรู้เพื่อพัฒนากฎที่ใช้ในการจำแนกกลุ่มข้อมูลของอินพุท โดยจะมี 2 ส่วนด้วยกัน ได้แก่ ส่วนการเรียนรู้แบบเสริมกำลัง ใช้ในการปรับค่าพารามิเตอร์ที่เกี่ยวข้องกับการจำแนกกลุ่มข้อมูล และกระบวนการค้นหากฎ (Discovery operator) เพื่อค้นหากฎที่มีประสิทธิภาพมากขึ้น

2.3.1 ส่วนการเรียนรู้แบบเสริมกำลัง (Reinforcement Learning Component)

การเรียนรู้ในส่วนของการเรียนรู้แบบเสริมกำลัง จะเกิดหลังจากได้รับคำตอบแทนจากสภาพแวดล้อม โดยจะใช้ในการปรับค่า พารามิเตอร์ p , ε และ F ของกฎที่อยู่ในเซตการกระทำ Widrow-Hoff delta rule [3] โดยเริ่มจากค่า prediction payoff p ดังสมการต่อไปนี้

$$p(t) = p(t-1) + \beta \times (R - p(t-1)) \quad (2.2)$$

โดยที่ β คืออัตราการเรียนรู้ (Learning rate) ซึ่งมีค่าอยู่ในช่วง 0 ถึง 1 และ R เป็นคำตอบแทนที่ได้รับจากสถานะแวดล้อม โดยจะมีค่าระหว่าง ค่าตอบแทนต่ำสุด และค่าตอบแทนสูงสุด ถัดจากนั้นจะทำการปรับปรุงค่าความผิดพลาดของกฎ ε ซึ่งจะปรับปรุงดังนี้

$$\varepsilon(t) = \varepsilon(t-1) + \beta \times (|R - p(t)| - \varepsilon(t-1)) \quad (2.3)$$

ในการปรับปรุงค่าฟิตเนส F มีรายละเอียดดังนี้ เริ่มจากหาค่าความถูกต้องของกฎ (accuracy: K) หลังจากได้ค่าความถูกต้อง K จะนำมาหาค่าเฉลี่ยของค่าความถูกต้องเรียกว่า ค่าความถูกต้องสัมพัทธ์ (relative accuracy: K') ซึ่งทั้งสองสามารถคำนวณได้จากสมการดังนี้

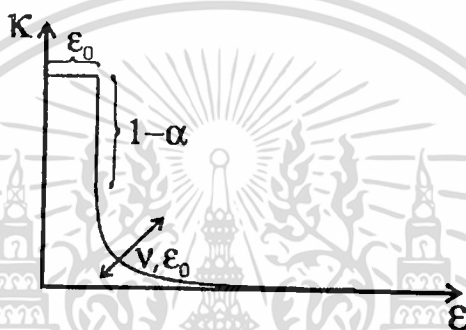
$$K = \begin{cases} 1, & \text{if } (\varepsilon < \varepsilon_0) \\ \alpha \left(\frac{\varepsilon}{\varepsilon_0} \right)^{-\nu}, & \text{otherwise} \end{cases} \quad (2.4)$$

$$K' = \frac{K}{\sum_{x \in [A]} K_x} \quad (2.5)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พารามิเตอร์ ε_0 เป็นความผิดพลาดที่ยอมรับได้ ที่ใช้ในการประเมินค่าความถูกต้องของกฎ โดยค่า fitness จะการประมาณค่าความผิดพลาดของกฎ โดยถ้าค่าความผิดพลาด ε ของกฎมีค่าน้อยกว่า ε_0 จะถือว่ายอมรับได้และกำหนดให้มีค่าความถูกต้องของกฎ K ของกฎนั้นเป็น 1 ค่า ε_0 นี้จะต้องกำหนดให้มีค่ามากกว่า 0

ในส่วนของพารามิเตอร์ α และพารามิเตอร์ ν ทั้งสองเป็นค่าที่คอยควบคุมค่าความถูกต้องของกฎ K (ในกรณีที่ค่าความผิดพลาด ε มีค่ามากกว่าค่า ε_0) ถ้าค่าความผิดพลาด ε ของกฎนั้นมีค่ามากจะส่งผลให้ค่าความถูกต้อง K ยิ่งน้อยลง โดยพารามิเตอร์ α จะมีค่าอยู่ในช่วง 0 ถึง 1 และพารามิเตอร์ ν จะมีค่ามากกว่า 0



รูป 2.4 กราฟการปรับค่าความถูกต้อง K

ซึ่งการคำนวณค่าความถูกต้อง K นั้นสามารถอธิบายได้จากรูป 2.4 สังเกตได้ว่าค่า ε_0 จะคอยควบคุมระดับการยอมรับความผิดพลาดที่เกิดขึ้น ถ้าค่าผิดพลาด ε น้อยกว่า ε_0 จะให้ค่า K สูงสุดและเท่ากัน ค่า α จะคอยควบคุมระดับความต่างของค่า K ระหว่างกฎที่ถูกต้องกับกฎที่ยังไม่ถูกต้อง และค่า ν จะคอยควบคุมความต่างของค่าผิดพลาดกฎที่เกิดขึ้น ถ้าค่า ν เพิ่มขึ้นจะเห็นความแตกต่างของค่าความผิดพลาดได้ชัดเจนขึ้น

หลังจากนั้นค่าความถูกต้อง K ของกฎที่อยู่ใน action set $[A]$ จะถูกเปลี่ยนเป็นค่าความถูกต้องสัมพัทธ์ (relative accuracy K') ตามสมการที่ 5 หลังจากได้ค่า K' แล้วก็จะทำการปรับปรุงค่าฟิตเนส F ดังนี้

$$F(t) = F(t-1) + \beta \times (K' - F(t-1)) \quad (2.6)$$

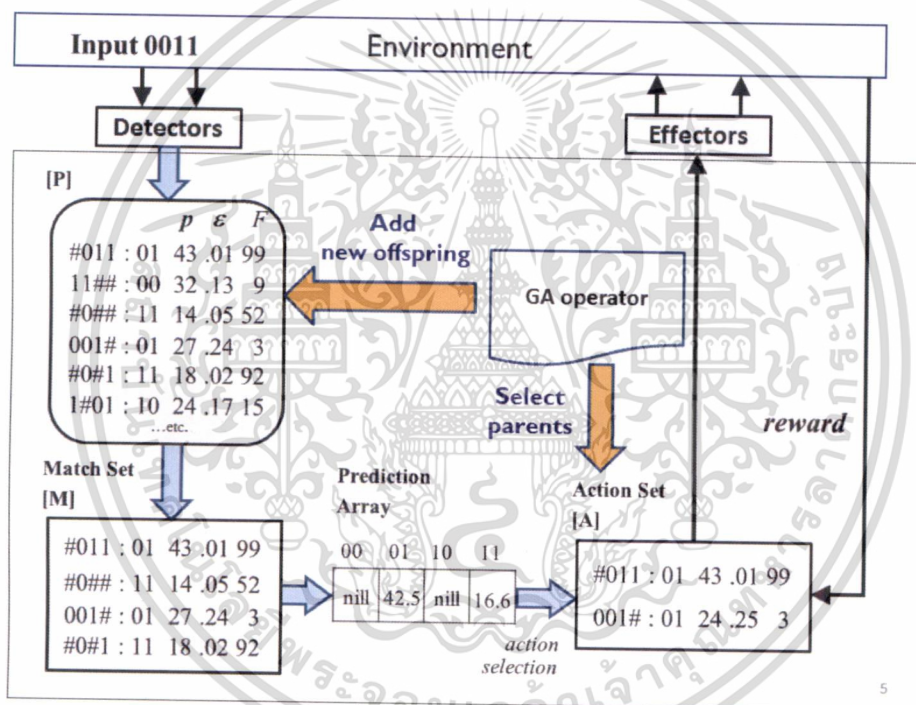
โดยค่า fitness ของกฎจะถูกปรับตามค่าความถูกต้องที่สัมพันธ์กับกฎอื่นๆ ในเซตการกระทำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2 กระบวนการค้นหาตัวจำแนกประเภท (Discovery Component)

ใน XCS ใช้วิธีการค้นหาตัวกฎ 2 วิธี วิธีแรกคือกระบวนการเจเนติกอัลกอริทึม [4] และกระบวนการสุ่มสร้างกฎที่เรียกว่า covering

กระบวนการเจเนติกอัลกอริทึมจะเกิดขึ้นเมื่อถึงช่วงเวลาที่กำหนดไว้ โดยจะพิจารณาจากรอบการทำงานปัจจุบันเทียบกับค่าเทียบกับค่าเฉลี่ยรอบการทำงานของเจเนติกครั้งล่าสุด time stamp t_s ของกฎในเซตการกระทำ ซึ่งถ้ามีค่ามากกว่าช่วงเวลาเวลาที่กำหนดไว้สำหรับการทำงานของเจเนติกอัลกอริทึมที่กำหนดไว้ (GA threshold) θ_{GA} ก็จะทำการเรียนรู้ใช้เจเนติกอัลกอริทึมกับกฎในเซตการกระทำ



รูป 2.5 การทำงานของเจเนติกอัลกอริทึมใน XCS

การทำงานของเจเนติกอัลกอริทึมใน XCS เป็นดังรูป 2.5 เจเนติกอัลกอริทึมจะเลือกกฎสองกฎจากเซตการกระทำ เพื่อนำมาเป็นพ่อแม่พันธุ์ โดยใช้กระบวนการแลกเปลี่ยนยีนส์ (Crossover) และการผ่าเหล่า (Mutation) ในการสร้างกฎใหม่ขึ้น หลังจากได้กฎใหม่แล้วจะนำมาแทนที่กฎเดิมที่มีอยู่ ซึ่งจะกล่าวถึงในหัวข้อถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนกระบวนการ covering จะเกิดขึ้นก็ต่อในเซตสอดคล้องมีความหลากหลายของ ส่วนการกระทำน้อยกว่าค่า Minimum action threshold θ_{mna} ซึ่งเป็นข้อกำหนดขั้นต่ำของความหลากหลายของการกระทำที่อยู่ในเซตสอดคล้อง กระบวนการ covering จะสุ่มสร้างกฎใหม่ขึ้นมาโดยสุ่มให้บาง บิตของอินพุต เป็น # และสุ่มค่าการกระทำที่ยังไม่มีในเซตสอดคล้องให้กฎใหม่ที่สร้างขึ้นมา

2.3.3 การลบกฎ (Classifier Deletion)

ถ้าจำนวนของกฎในเซตประชากรทั้งหมด มีค่ามากเกินกว่าจำนวนประชากรสูงสุดที่กำหนดไว้ N ตัว ระบบจำแนกประเภทแบบเรียนรู้ได้โดยอาศัยค่าความถูกต้องจะลบกฎบางกฎออกจากเซตประชากรให้เหลือ N ตัว XCS จะพิจารณาหากฎที่จะ โดยลบจากกฎทั้งหมดที่มีอยู่ในเซตประชากร โดยจะพิจารณาจากค่าประมาณขนาดของเซตการกระทำที่กฎนั้นเคยปรากฏอยู่ ถ้ามีมากก็จะมีโอกาสโดนลบมาก และจะพิจารณาค่าประสิทธิภาพ exp ของกฎที่จะเพิ่มขึ้นทุกครั้งเมื่อกฎนั้นเข้าไปอยู่ในเซตการกระทำ ถ้ากฎนั้นมีประสิทธิภาพมากแต่มีค่าฟิตเนส F น้อยก็จะมีโอกาสให้ถูกลบมากยิ่งขึ้น

2.3.4 Microclassifiers

XCS มีวิธีการเพิ่มความเร็วในการค้นหากฎที่สอดคล้องกับอินพุตที่เข้ามาจากเซตประชากร [P] โดยให้ชื่อว่า macroclassifier วิธีการเพิ่มความเร็วด้วย macroclassifier จะนำกฎที่มีส่วนของเงื่อนไขและการกระทำเหมือนกันจะถูกนำมารวมกัน และเพิ่มพารามิเตอร์ที่เรียกว่า *numerosity* เพื่อใช้นับจำนวนกฎที่ซ้ำซ้อนกันให้เป็นกฎเดียว โดยวิธี macroclassifier จะลดเวลาในการเปรียบเทียบกฎลง เพราะวิธี macroclassifier ลดความซ้ำซ้อนของกฎลง

ในการลบกฎจะต้องพิจารณาที่ค่า *numerosity* ก่อนว่ามีค่ามากกว่า 1 หรือไม่ ถ้ามากกว่าก็จะทำการลดค่า *numerosity* ลง 1 โดยที่จะไม่ลบกฎนั้น จนกระทั่งค่า *numerosity* มีค่าเหลือ 1 จึงทำการลบกฎนั้นออกจากเซตประชากร [P] ได้ รวมถึงการกระทำต่างๆ กับกฎในเซตประชากรก็จะคิดในเชิงของ macroclassifier ด้วย เช่นในการคิดจำนวนกฎทั้งหมดในเซตประชากรก็จะคิดจากค่าผลรวมของค่า *numerosity* ของกฎในเซตประชากร ได้ เนื่องจาก *numerosity* เป็นที่แสดงว่ามีกฎซ้ำซ้อนกันเท่าไร เมื่อค่า *numerosity* ลดลงหรือเพิ่มขึ้น ก็แปลว่ามีการลบกฎหรือเพิ่มกฎ ได้เช่นเดียวกัน

2.3.5 กระบวนการตรวจวัดความซ้ำซ้อนของกฎ (Subsumption Detection)

ใน XCS จะมีการตรวจวัดความซ้ำซ้อนของกฎเพื่อลดเวลาในการค้นหากฎ และเพิ่มความคงทนของการเรียนรู้ที่ได้ การตรวจวัดความซ้ำซ้อนของกฎจะมีการตรวจวัด 2 แบบ คือ การตรวจวัดความซ้ำซ้อนที่เกิดจากระบวนการเจเนติกอัลกอริทึม (GA subsumption Detection) และการตรวจวัดความซ้ำซ้อนของกฎที่อยู่ในเซตการกระทำ (action set subsumption Detection)

การตรวจวัดความซ้ำซ้อนที่เกิดจากระบวนการเจเนติกอัลกอริทึม จะตรวจวัดว่ากฎใหม่ที่ได้จากระบวนการเจเนติกอัลกอริทึมมีส่วนของเงื่อนไขซ้ำซ้อนกับพ่อแม่หรือไม่ ถ้ากฎใหม่ที่ได้ซ้ำซ้อนกับพ่อแม่ก็จะทำการเพิ่มค่า *numerosity* ของพ่อแม่ที่มีลูกซ้ำซ้อนนั้น โดยไม่เพิ่มกฎที่ได้ใหม่นี้เข้าไปในเซตประชากร

การตรวจวัดความซ้ำซ้อนที่อยู่ในเซตการกระทำ จะตรวจวัดว่าในเซตการกระทำมีกฎที่ความซ้ำซ้อนกันอยู่หรือไม่ โดยจะหากฎที่มีความครอบคลุมมากที่สุด โดยที่กฎนั้นต้องมีค่าความถูกต้องและค่าประสิทธิผลมากพอ จากนั้นทุกๆ กฎที่อยู่ในเซตจะทำการทดสอบว่าซ้ำซ้อนกับกฎนั้นหรือไม่ ถ้าซ้ำซ้อนก็จะกำจัดออกจากเซตประชากรและเพิ่มค่า *numerosity* ให้กับกฎที่มีความครอบคลุมมากที่สุด

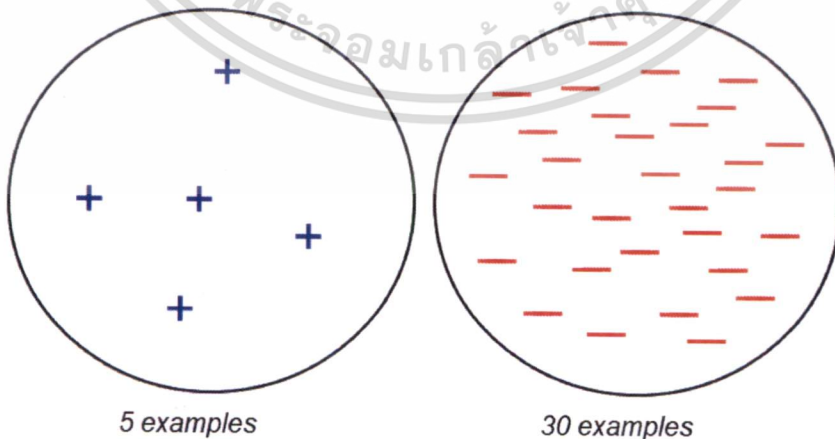
บทที่ 3

ชุดข้อมูลที่ไม่สมดุล

และแนวทางการแก้ไขปัญหาความไม่สมดุลของชุดข้อมูล

3.1 ชุดข้อมูลที่ไม่สมดุล

ชุดข้อมูลที่ไม่สมดุล (Imbalance Dataset) คือ ชุดข้อมูลที่มีจำนวนของข้อมูลในกลุ่มใดกลุ่มหนึ่งมากกว่าจำนวนข้อมูลในกลุ่มอื่นมาก ๆ ดังรูป 3.1 ความแตกต่างของจำนวนข้อมูลนี้ถ้ามีมากก็ จะส่งผลกระทบต่อการเรียนรู้ของจักรกลเรียนรู้ เนื่องจากจักรกลเรียนรู้โดยส่วนใหญ่แล้วพัฒนาขึ้นมาโดยมีสมมติฐานว่า “จำนวนข้อมูลแต่ละกลุ่มในชุดข้อมูลฝึกสอน (data training) มีจำนวนเท่าๆ กันในแต่ละกลุ่ม” [6] ซึ่งเรียกว่าได้ว่ามีความสมดุลกันระหว่างกลุ่มข้อมูล (balanced class) เมื่อนำ จักรกลเรียนรู้เหล่านี้มาใช้กับชุดข้อมูลที่ไม่สมดุลแล้ว การเรียนรู้ของจักรกลเรียนรู้เหล่านี้จะโน้ม เียงเข้าหากลุ่มข้อมูลที่มีข้อมูลจำนวนมาก ข้อมูลที่เก็บได้จากสถานการณ์จริง (real data) โดยส่วน ใหญ่จะมีลักษณะไม่สมดุล เช่น การตรวจจับการบุกรุก (Intrusion detection) [6,7] และ การค้นหา แหล่งน้ำมันจากภาพถ่ายดาวเทียม (Oil spill in satellite image) [6] เป็นต้น ในปัจจุบันมีงานวิจัย หลายฉบับที่พยายามทดสอบจักรกลเรียนรู้กับปัญหาความไม่สมดุลของชุดข้อมูล เช่น งานวิจัยของ Nitesh V. Chawla ในบทความชื่อ “C4.5 and Imbalanced Data sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure” [8] และงานวิจัยของ Rehan Akbani และคณะ ในบทความชื่อ “Applying Support Vector Machines to Imbalanced Datasets” [9]



รูป 3.1 ตัวอย่างชุดข้อมูลที่ไม่สมดุล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 ดัชนีชี้วัดความไม่สมดุล (Degree of Imbalance)

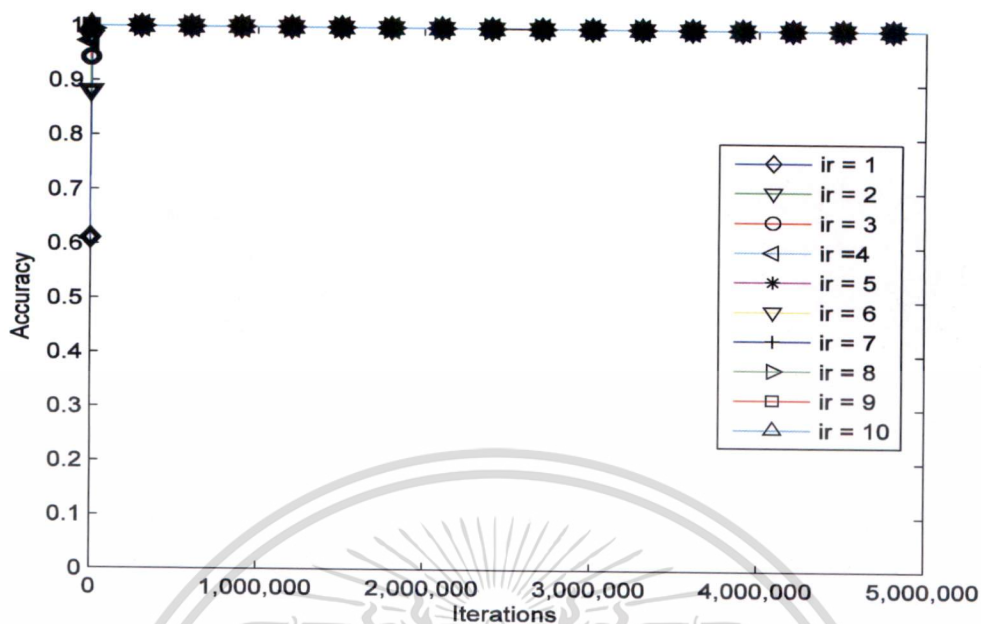
ดัชนีชี้วัดความไม่สมดุลเป็นตัวชี้วัดความสัมพันธ์ระหว่างกลุ่มข้อมูลที่มีจำนวนมากและกลุ่มข้อมูลที่มีจำนวนน้อย โดยแสดงในรูปของอัตราส่วนของจำนวนข้อมูลของกลุ่มข้อมูลที่มีจำนวนมากต่อจำนวนข้อมูลของกลุ่มข้อมูลที่มีจำนวนน้อย อัตราส่วนระหว่างจำนวนของทั้งสองกลุ่มนี้เรียกว่า อัตราส่วนความไม่สมดุล (imbalance ratio ir) [6] ในปี 2002 Nathalie Japkowicz และ Shaju Stephen [10] ได้นำเสนอผลงานวิจัยเกี่ยวกับการศึกษาความสัมพันธ์ระหว่าง ความไม่สมดุลของชุดข้อมูล และประสิทธิภาพในการจำแนกกลุ่มข้อมูลของต้นไม้การไม่ตัดสินใจ (decision tree) จากผลการศึกษาแสดงให้เห็นว่าอัตราส่วนของความไม่สมดุลที่มากขึ้นมีผลทำให้ประสิทธิภาพในการจำแนกกลุ่มข้อมูลของข้อมูลของจักรกลเรียนรู้ลดลง

3.3 การศึกษาผลกระทบของความไม่สมดุลของชุดข้อมูลต่อการเรียนรู้ของ XCS

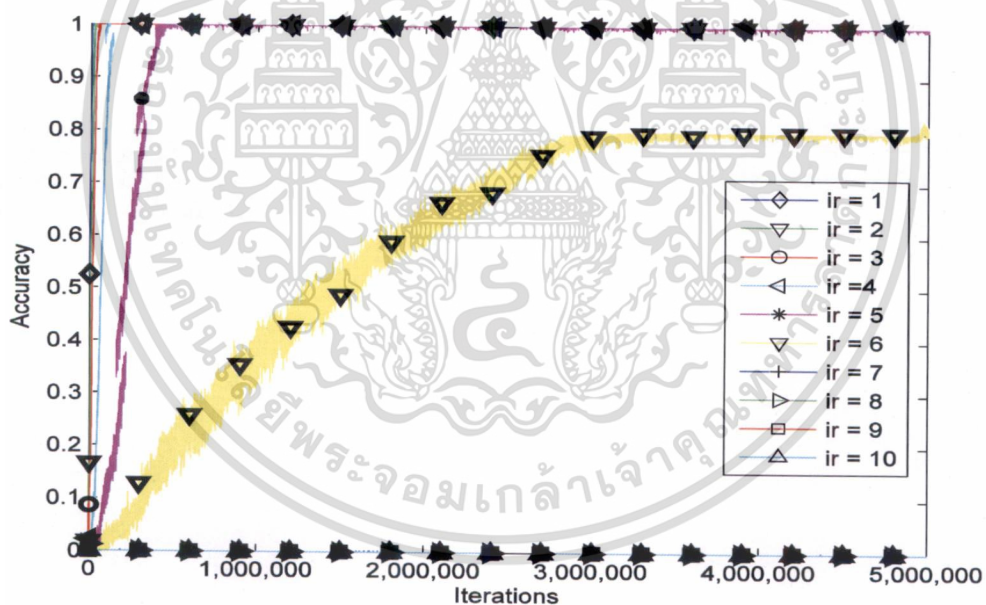
A. P. Orriols และ E. M. Bernardó [11] ได้ทำการศึกษาการทำงานของ XCS มาตรฐานกับชุดข้อมูลอุปกรณ์ผสมสัญญาณ (Multiplexer problem) ซึ่งเป็นชุดข้อมูลที่นิยมใช้ในการทดสอบเพื่อพัฒนา LCS ในทดลองนี้มีการกำหนดระดับความไม่สมดุลต่างกัน 10 ระดับ ตั้งแต่ระดับสมดุล $i = 1$ ซึ่งอัตราส่วนระหว่างจำนวนข้อมูลของกลุ่มข้อมูล 0 และกลุ่มข้อมูล 1 เป็น 2 เท่า และเมื่อระดับความไม่สมดุลเพิ่มขึ้น อัตราส่วนระหว่างจำนวนข้อมูลของกลุ่มข้อมูล 0 และกลุ่มข้อมูล 1 จะเพิ่มขึ้นระดับละ 2 เท่า ไปจนถึงระดับไม่สมดุล $i = 10$ ที่มีอัตราส่วนความไม่สมดุลเป็น 1024 ในการทดลองมีการกำหนดค่าพารามิเตอร์ต่างๆ ดังนี้

$$\text{Population size } (N) = 800, \beta = 0.2, \alpha = 0.1, \varepsilon_0 = 1, \nu = 5, \theta_{G_A} = 25, \\ \chi = 0.8, \mu = 0.04, \theta_{del} = 25, \delta = 0.1, \theta_{sub} = 200 \text{ และ } P_{\#} = 0.6$$

เพื่อหลีกเลี่ยงผลกระทบจากการทำ generalization จึงไม่ใช้ การตรวจสอบความซ้ำซ้อนบนเซตการกระทำ และใช้การตรวจวัดความซ้ำซ้อนที่เกิดจากกระบวนการเจเนติกอัลกอริทึมกับกฎที่ผ่านการเรียนรู้มานานมากๆ โดยการตั้งค่า θ_{sub} เป็น 200 โดยมีผลลัพธ์ดังต่อไปนี้



ก) True negative rate



ข) True positive rate

รูป 3.2 ผลการทดสอบ XCS มาตรฐาน กับชุดข้อมูลปัญหาหมัดติเฟล็กซ์เซอร์ 11 บิตที่ระดับความไม่สมดุลต่างกัน 10 ระดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูป 3.2 แสดงค่า true negative rate (TN) คืออัตราส่วนความถูกต้องจากการจำแนกกลุ่มข้อมูล '0' ต่อจำนวนข้อมูลในกลุ่ม '0' ทั้งหมด และค่า true positive rate (TP) คืออัตราส่วนความถูกต้องจากการจำแนกกลุ่มข้อมูล '1' ต่อจำนวนข้อมูลในกลุ่ม '1' ทั้งหมด จากการทดสอบบนชุดข้อมูลปัญหาหมัดติเฟล็กซ์เซอร์ 11 บิต ที่กำหนดระดับความไม่สมดุลต่างๆ กัน 10 ระดับตั้งแต่ระดับที่สมดุลที่ $i=1$ ซึ่งอัตราส่วนความไม่สมดุลเป็น 2 ไปจนถึงระดับที่ไม่สมดุลระดับที่ 10 รูป 3.2 (ก) แสดงให้เห็นว่า XCS มาตรฐานสามารถเรียนรู้กลุ่ม '0' ซึ่งเป็นกลุ่มข้อมูลที่มีจำนวนมากได้อย่างถูกต้องและรวดเร็ว โดยใช้จำนวนรอบของการเรียนรู้เพียง 10,000 รอบ แต่ทว่าที่รูป 3.2 (ข) แสดงให้เห็นว่า XCS มาตรฐานสามารถเรียนรู้ลักษณะของกลุ่มข้อมูล '1' ซึ่งเป็นกลุ่มข้อมูลที่มีจำนวนน้อยได้อย่างถูกต้องจนตั้งแต่ระดับความไม่สมดุลที่ 1 ไปจนถึงระดับความไม่สมดุลที่ 5 หลังจากผ่านการเรียนรู้ไปแล้ว 10,000, 20,000, 40,000, 80,000 และ 300,000 รอบ ตามลำดับ ในขณะที่กราฟการเรียนรู้ของระดับความไม่สมดุลที่ 6 คงที่อยู่ที่ย่อละ 80 หลังจากผ่านการเรียนรู้ไปแล้ว 250,000 รอบ และ XCS มาตรฐานไม่สามารถเรียนรู้กลุ่มข้อมูล '1' ได้เลยตั้งแต่ระดับความไม่สมดุลที่ 7 ไปจนถึงระดับที่ 10

จากการทดสอบ XCS มาตรฐานกับชุดข้อมูลหมัดติเฟล็กซ์เซอร์ 11 บิต พบว่า XCS มาตรฐานสามารถเรียนรู้ชุดข้อมูลที่ไม่สมดุลได้อย่างถูกต้องจนกระทั่งระดับความไม่สมดุลเพิ่มขึ้นถึงระดับที่ $i = 5$ ที่อัตราส่วนความไม่สมดุลเป็น $2^5 = 32$ และเมื่อระดับความไม่สมดุลของชุดข้อมูลสูงขึ้นเกินระดับที่ $i = 5$ แล้วความสามารถในการเรียนรู้ของ XCS มาตรฐานจะลดลง และเมื่อระดับความไม่สมดุลของข้อมูลสูงถึงระดับที่ $i = 7$ เป็นต้นไป XCS มาตรฐานไม่สามารถจำแนกข้อมูลในกลุ่มข้อมูลที่มีจำนวนข้อมูลน้อยได้เลย ดังรูป 3.2

ตาราง 3.1 รายละเอียดกฎของ XCS หลังจากฝึกสอนกับชุดข้อมูลหมัดติเฟล็กซ์เซอร์ที่ระดับความไม่สมดุลระดับที่ $i = 7$

Condition	Action	Prediction	Prediction Error	Fitness	Num.
#####	0	1000	1.2×10^{-4}	0.98	385
#####	1	1.2×10^{-4}	7.4×10^{-5}	0.98	366
...					

จากการทดสอบระบบ XCS มาตรฐานกับชุดข้อมูลหมัดติเฟล็กซ์เซอร์ 11 บิต เมื่อพิจารณากฎที่ได้หลังจากการเรียนรู้ชุดข้อมูลหมัดติเฟล็กซ์เซอร์พบว่า มีกฎที่มีความเป็น overgeneral เกิดขึ้นเป็นจำนวนมาก และกฎที่มีความเป็น overgeneral เหล่านี้โดยส่วนใหญ่มีการเรียนรู้โน้มเอียงไปทางกลุ่มข้อมูล '0' ซึ่งเป็นกลุ่มข้อมูลที่มีข้อมูลเป็นจำนวนมาก ดังตาราง 3.1 มีกฎที่เป็น overgeneral

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

เป็นจำนวนมาก คือ มีกฎที่มีส่วนเงื่อนไข-การกระทำเป็น #####-0 จำนวน 385 ตัว และกฎที่มีส่วนเงื่อนไข-การกระทำเป็น #####-1 366 ตัว คิดเป็นร้อยละ $(385+366)*100/800 = 93.88$ ของ classifier ทั้งหมดที่มีอยู่ใน population set

สรุปได้ว่า XCS มาตรฐานสามารถจำแนกกลุ่มข้อมูลที่มีอัตราส่วนความไม่สมดุลในระดับที่ไม่สูงมากได้ แต่เมื่ออัตราส่วนความไม่สมดุลเกินกว่า 64/1 เท่า XCS มาตรฐาน จะลดลงและจำไม่สามารถเรียนรู้และจำแนกกลุ่มข้อมูลที่มีจำนวนน้อยได้เลยเมื่ออัตราส่วนความไม่สมดุลมีค่าตั้งแต่ 128/1 เท่าเป็นต้นไป โดยสาเหตุที่ทำให้ XCS มาตรฐานไม่สามารถจำแนกกลุ่มข้อมูลที่มีจำนวนน้อยได้เมื่ออัตราส่วนความไม่สมดุลมีค่ามาก คือมีกฎที่มีลักษณะเป็น over-general เกิดขึ้นเป็นจำนวนมาก โดยกฎเหล่านี้มีค่า Fitness ซึ่งเป็นค่าความน่าเชื่อถือของกฎที่สูงมาก และองค์ความรู้ที่ได้จากกฎเป็น over-general นี้โน้มเอียงเข้าหาข้อมูลที่มีจำนวนข้อมูลมากๆ

3.4 ความสัมพันธ์ของการกำหนดค่าตอบแทนสูงสุดค่าตอบแทนต่ำสุดและค่าความผิดพลาดที่ยอมรับได้ กับอัตราส่วนความไม่สมดุลของชุดข้อมูล

A. P. Orriols และ E. M. Bernardó [11] เสนอการวิเคราะห์ความสัมพันธ์ของค่าตอบแทนสูงสุด R_{max} ค่าตอบแทนสูงสุดต่ำสุด R_{min} และค่า error threshold ϵ_0 ต่อประสิทธิภาพในการจำแนกกลุ่มข้อมูลของ XCS ดังต่อไปนี้ โดยมีรายละเอียดดังต่อไปนี้

เนื่องจากการปรับค่าพารามิเตอร์ perdition payoff ของกฎใดๆ เป็นการปรับค่าโดยใช้ Widrow-Hoff delta rule ในการประมาณหาค่าเฉลี่ยที่ค่าตอบแทนที่ได้รับจากสภาพแวดล้อม ดังนั้นจึงประมาณได้ว่า

$$p_{cl} = \sum P(S)P(A|S)R(S, A_{cl}) \quad (3.1)$$

โดยที่ S คือ state หรืออินพุต XCS ได้รับจากสภาพแวดล้อม

cl คือ กฎหรือตัวจำแนกประเภทที่สอดคล้องกับอินพุต S

p_{cl} คือ ค่า perdition payoff ของกฎ cl

$P(S, A)$ คือ ค่าความน่าจะเป็นที่จะอินพุต S จะมาจากกลุ่มข้อมูล A

$P(S)$ คือ ค่าความน่าจะเป็นที่ XCS จะได้รับอินพุต S

$P(A|S)$ คือ ค่าความน่าจะเป็นที่ action A จะถูกเลือกเมื่อ XCS ได้รับอินพุต S

$R(S, A_{cl})$ คือ ตอบแทนจากที่ XCS จะได้รับจากการเลือก action A

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และใช้ Widrow-Hoff deltarule สำหรับปรับค่า prediction error ในการประมาณค่าความผิดพลาดที่เกิดขึ้นจากการจำแนกกลุ่มข้อมูลของกฎแต่ละกฎ

$$\varepsilon_{cl} = \sum P(S)P(A|S) |R(S, A_{cl}) - p_{cl}| \quad (3.2)$$

โดยที่ ε_{cl} คือ ค่า prediction error ของกฎ cl

คำนวณหาค่า prediction payoff และค่า prediction error ของกฎ ##### : 1 จะได้ว่า $P(S)$ มีค่าเป็น 1 เนื่องจาก กฎ ##### : 1 สามารถสอดคล้องกับข้อมูลทุกรูปแบบ

กำหนดให้ $P_{cl}(C)$ คือค่าความน่าจะเป็นที่กฎ cl จะจำแนกได้อย่างถูกต้อง ดังนั้นค่าความน่าจะเป็นที่กฎ cl จะจำแนกผิดพลาดมีค่าเป็น $1-P_{cl}(C)$ ในการจำแนกกลุ่มข้อมูล XCS จะได้รับค่าตอบแทนสูงสุดเมื่อจำแนกกลุ่มข้อมูลได้ถูกต้อง และจะได้รับค่าตอบแทนต่ำสุดเมื่อจำแนกกลุ่มข้อมูลผิดพลาด ซึ่งสามารถแทนค่าในสมการที่ 3.1 ได้ดังต่อไปนี้

$$p_{cl} = P_{cl}(C) \times R_{\max} + (1 - P_{cl}(C)) \times R_{\min} \quad (3.3)$$

$$\varepsilon_{cl} = P_{cl}(C) \times |R_{\max} - p_{cl}| + (1 - P_{cl}(C)) \times |R_{\min} - p_{cl}| \quad (3.4)$$

โดยทั่วไปแล้วการกำหนดค่าตอบแทนต่ำสุดในการใช้ XCS จำแนกกลุ่มข้อมูลจะกำหนดเป็น 0 ดังนั้นเมื่อแทนค่าตอบแทนต่ำสุดในสมการที่ 3.3 และ 3.4 ด้วยค่า 0 จะได้ว่า

$$p_{cl} = P_{cl}(C) \times R_{\max} \quad (3.5)$$

$$\varepsilon_{cl} = P_{cl}(C) \times |R_{\max} - p_{cl}| + (1 - P_{cl}(C)) \times p_{cl} \quad (3.6)$$

แทนค่า p_{cl} จากสมการที่ 3.3 ในสมการที่ 3.4 จะได้ว่า

$$\varepsilon_{cl} = P_{cl}(C) \times |R_{\max} - P_{cl}(C) \times R_{\max}| + (1 - P_{cl}(C)) \times P_{cl}(C) \times R_{\max} \quad (3.7)$$

$$\varepsilon_{cl} = P_{cl}(C) \times R_{\max} (1 - P_{cl}(C)) + (1 - P_{cl}(C)) \times P_{cl}(C) \times R_{\max} \quad (3.8)$$

$$\varepsilon_{cl} = P_{cl}(C) \times R_{\max} (2 - 2P_{cl}(C)) \quad (3.9)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\varepsilon_{cl} = 2 \times R_{\max} (P_{cl}(C) - P_{cl}(C)^2) \quad (3.10)$$

ค่า $P_{cl}(C)$ สามารถประมาณได้จำนวนครั้งในการจำแนกถูกต้องเทียบกับจำนวนครั้งในการจำแนกทั้งหมด ดังต่อไปนี้

$$P_{cl}(C) = \frac{C}{C+W} \quad (3.11)$$

โดยที่ C คือจำนวนครั้งในการจำแนกถูกต้อง และ W คือจำนวนครั้งในการจำแนกผิดพลาด

สำหรับกำหนดให้ i เป็นอัตราส่วนระหว่างจำนวนครั้งในการจำแนกถูกต้องและจำนวนครั้งในการจำแนกผิดพลาดได้ จะได้ว่า

$$i = \frac{W}{C} \quad (3.12)$$

แทนค่า i จากสมการ 3.12 ลงในสมการการหาค่า $P_{cl}(C)$ ในสมการ 3.11 จะได้ว่า

$$P_{cl}(C) = \frac{1}{1+i} \quad (3.13)$$

และแทนค่า $P_{cl}(C)$ จากสมการ 3.13 ลงในการหาค่า prediction error ในสมการ 3.11 จะได้

$$\varepsilon_{cl} = 2 \times R_{\max} \left(\frac{1}{1+i} - \frac{1}{(1+i)^2} \right) \quad (3.14)$$

$$\varepsilon_{cl} = 2 \times R_{\max} \left(\frac{1+i}{(1+i)^2} - \frac{1}{(1+i)^2} \right) \quad (3.15)$$

$$\varepsilon_{cl} = 2 \times R_{\max} \left(\frac{i}{(1+i)^2} \right) \quad (3.16)$$

สำหรับการคำนวณค่า accuracy ของแต่ละกฎจะพิจารณาเปรียบเทียบกับค่า ε_0 โดยกฎที่มีลักษณะเป็น overgeneral เนื่องจากกฎที่เป็น overgeneral จะครอบคลุมกลุ่มข้อมูลมากกว่า 1 กลุ่ม ดังนั้นการทำงานของ XCS ควรจะสามารถจำแนกกลุ่มข้อมูลที่มีจำนวนน้อยได้ ก็ต่อเมื่อ prediction error ε ของกฎที่มีลักษณะเป็น overgeneral มีค่ามากกว่า ε_0 ซึ่งสามารถสรุปได้ว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\varepsilon_{cl} > \varepsilon_0 \quad (3.17)$$

$$2 \times R_{\max} \left(\frac{i}{(1+i)^2} \right) > \varepsilon_0 \quad (3.18)$$

โดยสามารถสรุปได้ว่า การกำหนดค่าตอบแทนสูงสุดและค่า error ที่ยอมรับได้ มีความสัมพันธ์ต่ออัตราส่วนความไม่สมดุลของชุดข้อมูลดังสมการต่อไปนี้

$$-\varepsilon_0 i^2 + 2i(R_{\max} - \varepsilon_0) - \varepsilon_0 > 0 \quad (3.19)$$

เนื่องจากในการทดลองของ A. P. Oriols และ E. M. Bernardó กำหนดค่าตอบแทนสูงสุด R_{\max} ไว้เป็น 1000 และค่าความผิดพลาดที่ยอมรับได้ ε_0 เป็น 1 เมื่อแทนค่าสูงสุดนี้ในสมการ 3.19 จะได้ว่า

$$1 < i < 1998 \quad (3.20)$$

สรุปได้ว่าเมื่อกำหนดค่าตอบแทนสูงสุด R_{\max} ไว้เป็น 1000 และค่าความผิดพลาดที่ยอมรับได้ ε_0 เป็น 1 แล้ว XCS ควรจะจำแนกกลุ่มข้อมูลของชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลได้ 1998 เท่า

3.5 ความสัมพันธ์ของการกำหนดค่าอัตราการเรียนรู้กับอัตราส่วนความไม่สมดุล

เมื่อทำการวิเคราะห์หาค่า prediction payoff p และค่า prediction error ε ที่ได้จากการทดลอง A. P. Oriols และ E. M. Bernardó เปรียบเทียบกับค่า prediction payoff p และค่า prediction error ε ที่คำนวณจากสมการ 3.3 และ 3.4 ของ XCS มาตรฐาน ที่ระดับความไม่สมดุลที่ 7 ซึ่งมีอัตราส่วนของจำนวนข้อมูลจากกลุ่มข้อมูล 0 ต่อกลุ่มข้อมูล 1 เป็น 128 เท่า ดังตารางต่อไปนี้

ตาราง 3.2 เปรียบเทียบค่า prediction payoff และค่า prediction error ที่ได้จากการทดลอง

Rule	$P_c(C)$	Experiment		Theoretical	
		p	ε	p	ε
##### : 0	128/129	1000	1.2×10^{-4}	992.24	15.38
##### : 1	1/129	1.2×10^{-4}	7.4×10^{-5}	7.75	15.38

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตาราง 3.2 แสดงให้เห็นว่าค่า prediction payoff p และ prediction error ε ที่ได้จากการทดลองมีค่าน้อยกว่าค่าที่คำนวณได้จากทฤษฎี ทำให้การเรียนรู้

ไม่ตรงกับค่าที่ได้จากทฤษฎีไม่ตรงกัน prediction payoff p , prediction error ε ซึ่งแสดงให้เห็นว่าอัตราการเรียนรู้ของมีผลต่อการเรียนรู้ของชุดข้อมูลที่ไม่สมดุล

อัตราการเรียนรู้เป็นปัจจัยหนึ่งที่มีผลกระทบต่อการเรียนรู้กับชุดข้อมูลที่ไม่สมดุลของ XCS เป็นอย่างมาก บทที่ 2 พบว่ากฎที่ได้หลังจากการเรียนรู้จะมีค่าของพารามิเตอร์ที่ใช้ในการทำนายต่างๆ โนม์เอียงเข้ากลุ่มที่มีจำนวนข้อมูลมาก ความโน้มเอียงนี้เกิดขึ้นเนื่องจากการใช้ Widrow-Hoff delta rule [5] ในการประมาณค่าของพารามิเตอร์ต่างๆ โดยไม่ได้สนใจต่ออัตราส่วนของความไม่สมดุลของจำนวนข้อมูลที่อยู่ในแต่ละกลุ่มข้อมูล การปรับค่าพารามิเตอร์ด้วยวิธี Widrow-Hoff delta rule นี้เป็นการหาค่าเฉลี่ยในช่วงเวลาหนึ่งซึ่งสามารถอธิบายได้ดังนี้

สมการทั่วไปในการหาค่าเฉลี่ยของค่า X จำนวน N ค่าคือ

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \quad (3.21)$$

ดังนั้นถ้าหาค่าเฉลี่ยของ X จำนวน $n+1$ ค่าคือ

$$\bar{X}_{n+1} = \frac{\sum_{i=1}^{n+1} X_i}{n+1} \quad (3.22)$$

$$\bar{X}_{n+1} = \frac{\sum_{i=1}^n X_i + X_{n+1}}{n+1} \quad (3.23)$$

กระจายออกเป็น 2 พจน์ดังนี้

$$\bar{X}_{n+1} = \frac{\sum_{i=1}^n X_i}{n+1} + \frac{X_{n+1}}{n+1} \quad (3.24)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากสมการ 3.24 $\frac{\sum_{i=1}^N X_i}{n} = \bar{X}_n$ ดังนั้น จะได้

$$\bar{X}_{n+1} = \frac{n\bar{X}_n + X_{n+1}}{n+1} \quad (3.25)$$

บวกและลบ $\bar{X}_n/(n+1)$ ในฝั่งขวาของสมการที่ 3.25 ดังนี้

$$\bar{X}_{n+1} = \frac{n\bar{X}_n + \bar{X}_n - \bar{X}_n + X_{n+1}}{n+1} \quad (3.26)$$

แยก $-\bar{X}_n/(n+1)$ ออกจากพจน์ที่ 1 ได้ดังนี้

$$\bar{X}_{n+1} = \frac{(n+1)\bar{X}_n}{n+1} + \frac{-\bar{X}_n + X_{n+1}}{n+1} \quad (3.27)$$

$$\bar{X}_{n+1} = \bar{X}_n + \frac{-\bar{X}_n + X_{n+1}}{n+1} \quad (3.28)$$

$$\bar{X}_{n+1} = \bar{X}_n + \frac{1}{n+1}(X_{n+1} - \bar{X}_n) \quad (3.29)$$

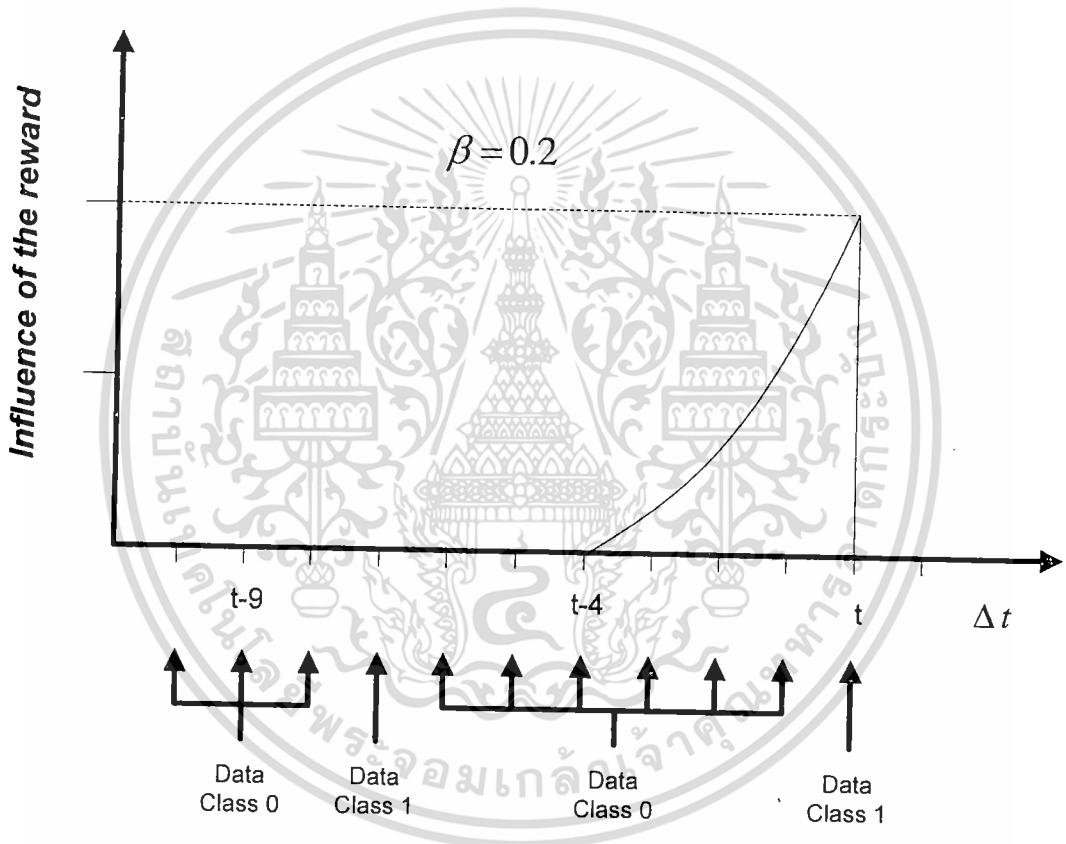
กำหนดให้ β มีค่าเป็น $\frac{1}{n+1}$ จะได้ว่า

$$\bar{X}_{n+1} = \bar{X}_n + \beta \cdot (X_{n+1} - \bar{X}_n) \quad (3.30)$$

สรุปได้ว่าสมการปรับค่าพารามิเตอร์ตามวิธีการ Widrow-Holf delta rule เป็นการหาค่าเฉลี่ยของค่าพารามิเตอร์ในช่วงเวลา $\frac{1}{\beta}$ หรือจาก ข้อมูลตัวล่าสุดที่ได้รับเข้ามา และข้อมูลที่ระบบจักรกลเรียนรู้เคยได้รับมาล่าสุด n ตัว ดังนั้นในการกำหนดค่าอัตราการเรียนรู้ของจักรกลเรียนรู้เพื่อประยุกต์ใช้กับข้อมูลที่ไม่สมดุลควรกำหนดค่าอัตราการเรียนรู้ให้สัมพันธ์กับอัตราส่วนความไม่สมดุลของข้อมูล โดยค่าอัตราการเรียนรู้ที่กำหนดควรมีค่าน้อยกว่าหรือเท่ากับค่าอัตราส่วนความไม่สมดุลของชุดข้อมูล ดังตัวอย่างต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างที่ 1 สมมติให้ใช้งาน XCS กับชุดข้อมูล Multiplexer 6 บิต ที่กำหนดค่าอัตราส่วนความไม่สมดุลของข้อมูลเป็น 6 ต่อ 1 โดยกำหนดให้กลุ่มข้อมูล 0 เป็นกลุ่มข้อมูลที่มีจำนวนมาก และกลุ่มข้อมูล 1 เป็นกลุ่มข้อมูลที่มีจำนวนน้อย และกำหนดให้ค่าอัตราการเรียนรู้ของ XCS มีค่าเป็น 0.2 ซึ่งสามารถคำนวณหาช่วงเวลาในการหาค่าเฉลี่ยได้ว่า $1/0.2 = 4+1$ iteration ดังนั้นในการเรียนรู้กับข้อมูลที่มีอัตราส่วนความไม่สมดุลเป็น 8 ต่อ 1 จึงทำให้ XCS ไม่สามารถเรียนรู้ของกลุ่มข้อมูลที่มีจำนวนน้อยได้อย่างถูกต้อง เนื่องจากการปรับค่าพารามิเตอร์ที่เกิดขึ้นนั้นเป็นการหาค่าเฉลี่ยจากในช่วงเวลา 5 iteration เท่านั้น เมื่อมีข้อมูลในกลุ่ม 1 เข้ามาระบบจะไม่นำค่าข้อมูลผลตอบแทนที่ได้รับจากข้อมูลจากกลุ่ม 1 ที่เข้ามาก่อนหน้านี้ได้ ดังตัวรูปต่อไปนี้

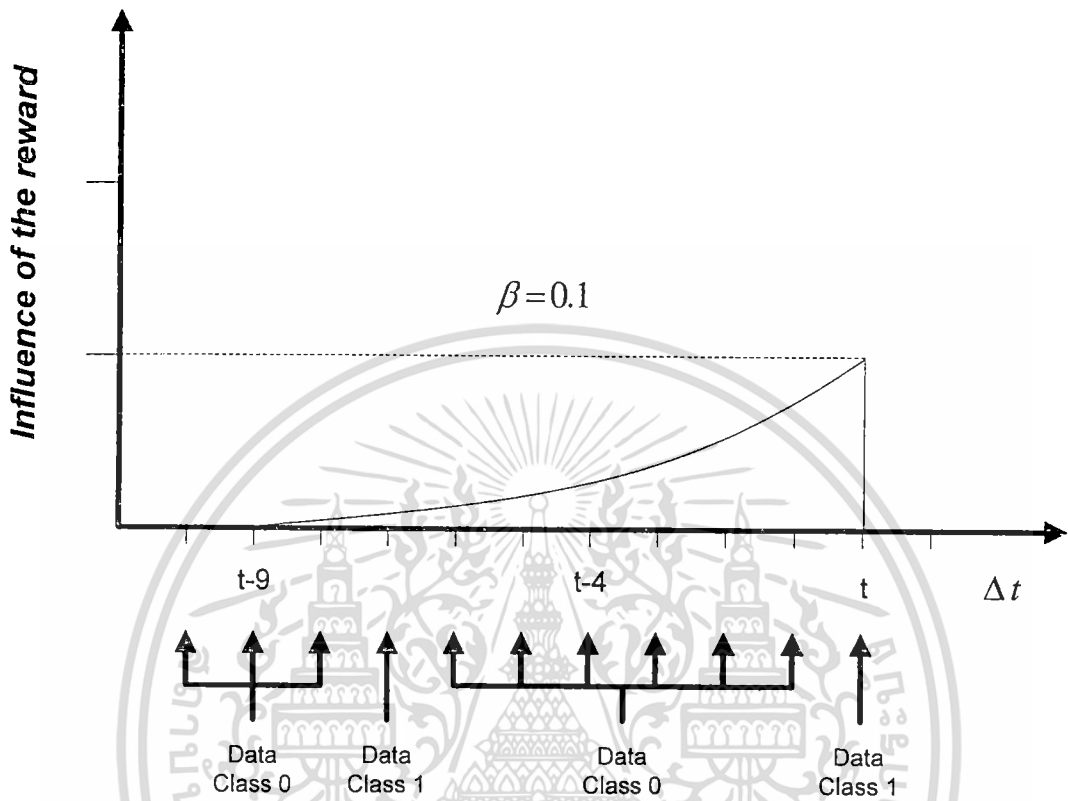


รูป 3.3 ช่วงเวลาในการเรียนรู้ของ XCS โดยกำหนดค่าอัตราการเรียนรู้เป็น 0.2

ตัวอย่างที่ 2 สมมติให้ใช้งาน XCS กับชุดข้อมูล Multiplexer 6 bit ที่กำหนดค่าอัตราส่วนความไม่สมดุลของข้อมูลเป็น 6 ต่อ 1 โดยกำหนดให้กลุ่มข้อมูล 0 เป็นกลุ่มข้อมูลที่มีจำนวนมาก และกลุ่มข้อมูล 1 เป็นกลุ่มข้อมูลที่มีจำนวนน้อย และกำหนดให้ค่าอัตราการเรียนรู้ของ XCS มีค่าเป็น 0.1 ซึ่งสามารถคำนวณหาช่วงเวลาในการหาค่าเฉลี่ยได้ว่า $1/0.1 = 9+1$ iteration ดังนั้นในการเรียนรู้กับข้อมูลที่มีอัตราส่วนความไม่สมดุลเป็น 6 ต่อ 1 ทำให้ XCS สามารถเรียนรู้ของกลุ่มข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่มีจำนวนน้อยได้อย่างถูกต้อง โดยระบบสามารถที่จะนำค่าข้อมูลผลตอบแทนที่ได้รับจากข้อมูลจากกลุ่ม 1 ที่เข้ามาก่อนหน้านี้ได้ ดังรูปต่อไปนี้



รูป 3.4 ช่วงเวลาในการเรียนรู้ของระบบ XCS โดยกำหนดค่าอัตราการเรียนรู้เป็น 0.1

จากการศึกษาผลกระทบของอัตราส่วนความไม่สมดุล ควรกำหนดค่าอัตราการเรียนรู้ให้มีค่าน้อยลงตามช่วงเวลาที่ได้รับข้อมูลจากกลุ่มข้อมูลที่มีจำนวนน้อยอย่างน้อย สมมุติให้ชุดข้อมูลมีอัตราส่วนความไม่สมดุลเป็น ir หมายความว่าโดยเฉลี่ยจะมีข้อมูลจากกลุ่มข้อมูลที่มีจำนวนมาก ir ข้อมูล ต่อข้อมูลจากกลุ่มข้อมูลที่มีจำนวนมาก 1 ข้อมูล จากการปรับค่าตัวแปรด้วยสมการ Widrow-Holf delta rule ซึ่งเป็นการหาค่าเฉลี่ยในช่วงเวลา n ข้อมูล โดยจะต้องกำหนดค่าอัตราการเรียนรู้ให้รับข้อมูลจากกลุ่มข้อมูล ที่มีจำนวนน้อยอย่างน้อย 1 ข้อมูลเสมอ ดังนั้นค่าช่วงเวลา n ข้อมูลจะมีค่าเท่ากับ $ir + 1$ ข้อมูล ดังนั้นการกำหนดค่าอัตราการเรียนรู้จึงควรกำหนดเป็น $1/(1+ir)$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

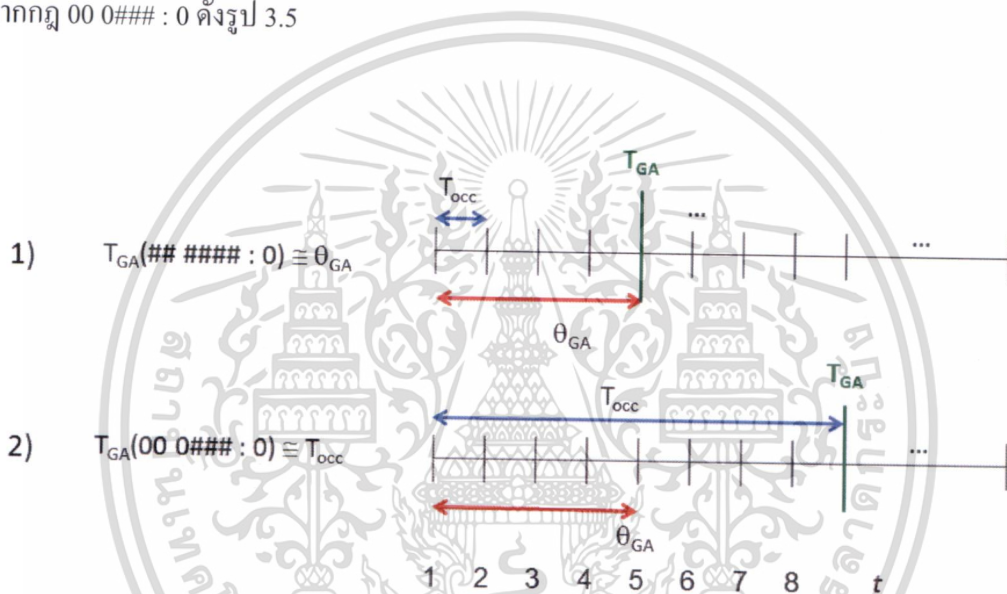
3.6 ความสัมพันธ์ของช่วงเวลาในการทำงานของเจเนติกอัลกอริทึม

ในหัวข้อนี้อธิบายถึงผลกระทบจากช่วงเวลาการทำงานของเจเนติกอัลกอริทึมที่มีผลต่อปัญหาการเกิด over-general จากหัวข้อที่ 3.3 ซึ่งกฎที่ได้จากการเรียนรู้เกือบทั้งหมดเป็นกฎที่มีลักษณะ over-general โดยปัจจัยหลักที่เกี่ยวกับการมีอยู่ของกฎที่มีลักษณะเป็น over-general เป็นจำนวนมากคือ กระบวนการสร้างกฎ โดย XCS มีอยู่สองกระบวนการคือ covering และเจเนติกอัลกอริทึม แต่ว่ากระบวนการ covering เป็นกระบวนการสร้างกฎด้วยการสุ่ม ดังนั้นปัจจัยที่ทำให้มีเกิดกฎที่มีลักษณะเป็น over-general จึงน่าจะเป็นผลกระทบจากการทำงานของเจเนติกอัลกอริทึม ต่อไปนี้จะวิเคราะห์ถึงสาเหตุและผลกระทบของกระบวนการเจเนติกอัลกอริทึมที่มีต่อการเรียนรู้ของ XCS กับชุดข้อมูลที่มีปัญหาความไม่สมดุลของข้อมูล

กระบวนการเจเนติกอัลกอริทึมเพื่อสร้างกฎใหม่ของ XCS จะเกิดขึ้นก็ต่อเมื่อ ค่าเฉลี่ยของช่วงเวลาที่ไม่มีการทำงานของเจเนติกอัลกอริทึม (นับจากครั้งล่าสุดจนถึงรอบปัจจุบัน) ของทุกๆ กฎที่อยู่ใน action set [A] มีค่าไม่น้อยกว่า ค่าช่วงเวลาขั้นต่ำในการทำงานของเจเนติกอัลกอริทึม θ_{GA} ที่กำหนดไว้ ดังนั้นเมื่อใช้งาน XCS ร่วมกับชุดข้อมูลที่ไม่สมดุลซึ่งมีจำนวนข้อมูลในแต่ละกลุ่มข้อมูลไม่เท่ากัน จะส่งผลให้มีการทำงานของกระบวนการเจเนติกอัลกอริทึมเกิดขึ้นกับกฎที่มีความสอดคล้องกับกลุ่มข้อมูลที่มีข้อมูลจำนวนมากบ่อยครั้งกว่า การทำงานของกระบวนการเจเนติกอัลกอริทึมกฎที่มีความสอดคล้องกับกลุ่มข้อมูลที่มีข้อมูลจำนวนน้อย ทำให้ในฐานข้อมูลของกฎมีจำนวนของกฎที่มีความสอดคล้องกับกลุ่มข้อมูลที่มีจำนวนมาก มากกว่าจำนวนของกฎที่มีความสอดคล้องกับกลุ่มข้อมูลที่มีจำนวนน้อย ดังตัวอย่างต่อไปนี้

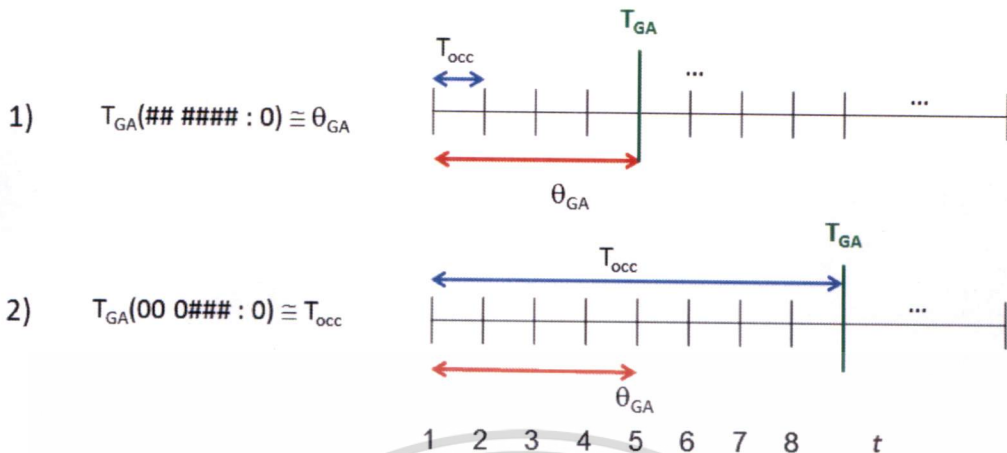
ตัวอย่างที่ 3 กำหนดให้ ชุดข้อมูล A เป็นชุดข้อมูลที่สร้างจากปัญหา Multiplexer 6 บิต และมีกฎอยู่ในระบบ 2 ลักษณะ ได้แก่ 1) ## ##### : 0 และ 2) 00 0### : 0 เมื่อพิจารณาที่จำนวนรูปแบบของอินพุตที่สอดคล้องกับกฎที่ ## ##### : 0 พบว่ามีจำนวนเป็น 2^6 ในขณะที่กฎ 00 0### : 0 สามารถสอดคล้องกับอินพุตได้จำนวน 2^3 รูปแบบ จะพบว่า จำนวนอินพุตที่กฎ ## ##### : 0 รองรับมีจำนวนเป็น 8 เท่าของจำนวนอินพุตที่กฎ 00 0### : 0 รองรับ ซึ่งถ้าเป็นชุดข้อมูลที่มีความสมดุลระหว่างจำนวนข้อมูลของกลุ่มข้อมูล 0 และกลุ่มข้อมูล 1 กฎ ## ##### : 0 จะถูกกำจัดออกจากระบบเนื่องจากมีค่า prediction error มากและมีค่า fitness น้อย แต่เมื่ออัตราส่วนความไม่สมดุลเพิ่มมากขึ้นทำให้ค่า prediction error และค่า fitness ของกฎ ## ##### : 0 ตามผลลัพธ์ที่ได้จากการทดลองในหัวข้อที่ 3.3 จะทำให้มีเกิดมีกฎ ## ##### : 0 เกิดขึ้นมาเป็นจำนวนมากถ้ามีการกำหนดค่าช่วงเวลาในการทำงานของเจเนติกอัลกอริทึมน้อยดังต่อไปนี้

สมมุติกำหนดให้ค่าช่วงเวลาในการทำงานของเจเนติกอัลกอริทึมมีค่าเท่ากับ 4 อินพุท ในขณะที่ช่วงเวลาเฉลี่ยที่กฎ $00\ 0### : 0$ จะได้เข้าไปในเซตการกระทำเป็น 8 อินพุท และช่วงเวลาเฉลี่ยที่กฎ $##\ ##### : 0$ จะได้เข้าไปในเซตการกระทำเป็น 1 อินพุท เมื่อพิจารณาที่ช่วงเวลาเฉลี่ยที่กฎ $00\ 0### : 0$ จะได้เป็นตัวเลือกของเจเนติกอัลกอริทึม ซึ่งมีค่าเป็น 8 อินพุทเท่ากับช่วงเวลาที่กฎ $00\ 0### : 0$ ได้เข้าไปในเซตการกระทำเปรียบเทียบกับ ช่วงเวลาเฉลี่ยที่กฎ $##\ ##### : 0$ จะได้เป็นตัวเลือกของเจเนติกอัลกอริทึม ซึ่งมีค่าเป็น 4 อินพุทเท่ากับช่วงเวลาที่ในการเรียกใช้เจเนติกอัลกอริทึมในการสร้างกฎ การกำหนดช่วงเวลาในการทำงานของเจเนติกอัลกอริทึมเป็น 4 อินพุทจะทำให้เกิดการสร้างกฎจากกฎ $##\ ##### : 0$ บ่อยครั้งกว่า กฎ $00\ 0### : 0$ เป็น 2 เท่าของการสร้างกฎจากกฎ $00\ 0### : 0$ ดังรูป 3.5



รูป 3.5 ช่วงเวลาในการทำงานของเจเนติกอัลกอริทึมกับกฎ $000### : 0$ และ $001### : 1$ ที่กำหนดช่วงเวลาในการใช้งานเจเนติกเป็น 4

จากตัวอย่างที่ 3 สมมุติกำหนดให้ค่าช่วงเวลาในการทำงานของเจเนติกอัลกอริทึมมีค่าเท่ากับ 8 อินพุท เท่ากับช่วงเวลาที่กฎ $00\ 0### : 0$ จะได้เข้าไปในเซตการกระทำดังรูป 3.6 จะทำให้กฎ $00\ 0### : 0$ และกฎ $##\ ##### : 0$ มีโอกาสได้เป็นตัวเลือกของเจเนติกอัลกอริทึมอย่างเท่าเทียมกัน



รูป 3.6 ช่วงเวลาในการทำงานของเจเนติกอัลกอริทึมกับกฎ 000#### : 0 และ 001#### : 1

สรุปได้ว่าการกำหนดค่าช่วงเวลาในการทำงานของเจเนติกอัลกอริทึมควรกำหนดตามช่วงเวลาที่กฎที่ครอบคลุมข้อมูลจากกลุ่มข้อมูลเพียงกลุ่มเดียว ซึ่งสามารถอนุมานได้ว่าเท่ากับช่วงเวลาที่ข้อมูลจากกลุ่มข้อมูลที่มีจำนวนน้อยเข้ามา เนื่องจากกฎที่ครอบคลุมข้อมูลจากกลุ่มข้อมูลที่มีจำนวนมาก มีโอกาสได้เข้าไปอยู่ในเซตการกระทำบ่อยครั้งกว่า กฎที่ครอบคลุมข้อมูลจากกลุ่มข้อมูลที่มีจำนวนน้อย

บทที่ 4

การปรับปรุง XCS สำหรับชุดข้อมูลที่ไม่สมดุล

4.1 บทนำ

ในบทนี้นำเสนอการปรับปรุง XCS สำหรับชุดข้อมูลที่ไม่สมดุล โดยคำนวณค่าอัตราส่วนความไม่สมดุลของชุดข้อมูลจากจำนวนครั้งที่จำแนกกลุ่มข้อมูลได้ถูกต้องและผิดพลาด การปรับค่าอัตราการเรียนรู้ และการขยายช่วงเวลาในการทำงานของเจเนติกอัลกอริทึม ตามอัตราส่วนความไม่สมดุลที่คำนวณได้ และนำเสนอการทดลองเพื่อวัดค่าประสิทธิภาพของ XCS กับชุดข้อมูลที่ไม่สมดุล ทั้งจากชุดข้อมูลที่สังเคราะห์ขึ้น และ ชุดข้อมูลจริงจาก UCI ซึ่งเป็นชุดข้อมูลมาตรฐานในการทดสอบจักรกลเรียน [12] และทำการเปรียบเทียบผลการทดลองที่ได้กับ XCS รูปแบบต่างๆ และเปรียบเทียบกับจักรกลเรียนรู้ชนิดอื่นๆ เช่น โครงข่ายประสาทเทียม (neural network), ต้นไม้ตัดสินใจ (Decision tree) ชนิด C4.5 และ SVM

4.2 การปรับปรุง XCS ให้สามารถเรียนรู้ชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลมากๆ ได้

ในหัวข้อนี้นำเสนอการคำนวณค่าอัตราส่วนความไม่สมดุลของชุดข้อมูล จากจำนวนครั้งที่จำแนกกลุ่มข้อมูลได้ถูกต้องและผิดพลาด การปรับค่าอัตราการเรียนรู้ และการขยายช่วงเวลาในการทำงานของเจเนติกอัลกอริทึม ตามอัตราส่วนความไม่สมดุลที่คำนวณได้ โดยอาศัยความสัมพันธ์ของการกำหนดค่าพารามิเตอร์ต่างๆ ของ XCS ที่ได้ศึกษาในบทที่ 3

4.2.1 การคำนวณอัตราส่วนความไม่สมดุลของชุดข้อมูล

เนื่องจากการเรียนรู้ของ XCS มีลักษณะเป็นการเรียนรู้แบบ reinforcement learning คือระบบจะรับข้อมูลสถานะเข้ามา และทำการตัดสินใจเลือกการกระทำหรือจำแนกกลุ่มข้อมูลส่งออกไปยัง สภาพแวดล้อม แล้วรอรับผลตอบแทนจากสภาพแวดล้อมกลับมา โดยที่ระบบจะได้เพียงว่า การกระทำที่เลือกหรือกลุ่มของข้อมูลที่จำแนกนั้น ถูกหรือผิด เท่านั้น ดังนั้นในการประมาณหาอัตราส่วนความไม่สมดุลจึงจำเป็นต้องคำนวณจากจำนวนครั้งที่ได้ค่าตอบแทนดีและไม่ดี ดังสมการต่อไปนี้

$$ir_{cl} = \begin{cases} 1 & \text{if } C_{cl} = 0, W_{cl} = 0 \\ C_{cl}/W_{cl} & \text{if } C_{cl} > W_{cl} \\ W_{cl}/C_{cl} & \text{if } C_{cl} \leq W_{cl} \end{cases} \quad (4.1)$$

โดยที่ cl คือ กฎหรือตัวจำแนกประเภทที่อยู่ในเซตการกระทำ

ir_{cl} เป็นค่าประมาณอัตราส่วนความไม่สมดุลของกฎ cl

C_{cl} คือจำนวนครั้งที่กฎ cl จำแนกถูกต้อง โดยพิจารณาจากค่าตอบแทนสูงสุด

W_{cl} คือจำนวนครั้งที่กฎ cl จำแนกผิดพลาด โดยพิจารณาจากค่าตอบแทนต่ำสุด

สมการ 4.1 พารามิเตอร์ C_{cl} และ W_{cl} จะเป็นตัวจำนวนครั้งที่ XCS ทำการเลือกการกระทำหรือกลุ่มของข้อมูลที่ถูกต้องหรือผิดพลาด โดยจะพิจารณาจากค่าตอบแทน ซึ่งถ้าได้ค่าตอบกลับมาเป็นค่าตอบแทนสูงสุดจะถือว่า XCS ทำการเลือกการกระทำหรือกลุ่มของข้อมูลได้ถูกต้อง แล้วทำการนับเพิ่มค่า C ของกฎ cl แต่ถ้าได้ค่าตอบแทนเป็น 0 จะถือว่า XCS ทำการเลือกการกระทำหรือกลุ่มของข้อมูลผิดพลาดจะทำการเพิ่มค่า W ของกฎ cl

และเนื่องจาก XCS มีการสร้างกฎขึ้นใหม่เรื่อยๆ โดยกฎที่สร้างใหม่เหล่านี้จะยังไม่สามารถคำนวณค่าอัตราส่วนความไม่สมดุลได้อย่างมีประสิทธิภาพ ดังนั้นจึงใช้การหาค่าเฉลี่ยของอัตราส่วนความไม่สมดุลจากกฎที่มีลักษณะใกล้เคียงกัน โดยกฎที่มีลักษณะใกล้เคียงกันนี้จะได้เข้าไปอยู่ในเซตการกระทำร่วมกันบ่อยครั้ง ซึ่งจะช่วยให้สามารถคำนวณค่าอัตราส่วนความไม่สมดุลได้เหมาะสมมากขึ้น แต่อย่างไรก็ตามการคำนวณค่าอัตราส่วนความไม่สมดุลจากจำนวนครั้งที่จำแนกถูกต้องและผิดพลาดมีการเปลี่ยนแปลงอย่างรวดเร็วในช่วงต้นๆ ของการเรียนรู้ ดังนั้นจึงใช้สมการ Widrow-Hoff delta rule ช่วยในการหาค่าเฉลี่ยที่แท้จริงของกฎที่อยู่ในเซตการกระทำดังต่อไปนี้

$$\psi_{cl}(t) = \psi_{cl}(t-1) + \beta \times (ir_{avg}^{-1} - \psi_{cl}(t-1)) \quad (4.2)$$

โดยที่ ir_{avg} เป็นค่าเฉลี่ยของอัตราส่วนความไม่สมดุลที่จากกฎใน action set

ψ_{cl} เป็นค่าถ่วงน้ำหนักการเรียนรู้ที่คำนวณได้อัตราส่วนความไม่สมดุลที่ปรากฏขึ้นในกฎ cl

4.2.2 การปรับค่าอัตราการเรียนรู้

ในหัวข้อนี้อธิบายการปรับค่าอัตราการเรียนรู้ของ XCS ตามความสัมพันธ์ระหว่างค่าอัตราส่วนความไม่สมดุล ต่อการกำหนดค่าอัตราการเรียนรู้ตามที่ได้อธิบายในหัวข้อที่ 3.3 ว่า อัตราการเรียนรู้ควรปรับให้สัมพันธ์กับช่วงเวลาที่ XCS ได้รับข้อมูลจากกลุ่มข้อมูลที่มีจำนวนน้อย

ใน XCS มีพารามิเตอร์ที่ใช้ในการเรียนรู้อยู่ 3 ตัว ได้แก่ prediction payoff p prediction error \mathcal{E} และ Fitness F พารามิเตอร์ที่ใช้ในการเรียนรู้ทั้ง 3 ตัวนี้มี 2 พารามิเตอร์ที่เกี่ยวข้องกับอัตราความไม่สมดุล 2 ตัวคือ prediction payoff p และ prediction error \mathcal{E} โดยพิจารณาจากสมการในการปรับค่าพารามิเตอร์ทั้ง 3 ในสมการ 2.2 2.3 และ 2.6 จะเห็นได้ว่าสมการ 2.2 และ 2.3 เป็นการปรับค่าจากค่าตอบแทนที่ได้รับ แต่สมการ 2.6 ซึ่งเป็นสมการปรับค่า Fitness จะปรับค่าจากค่าเฉลี่ยของความถูกต้องของกฎแทน ดังนั้นถ้าสามารถทำให้ค่าประมาณความผิดพลาดในการทำนายค่าตอบแทนสามารถรองรับอัตราส่วนความไม่สมดุลได้แล้วจะทำให้ค่าความแข็งแรงของกฎสามารถรองรับอัตราส่วนความไม่สมดุลได้เช่นเดียวกัน และสำหรับค่าประมาณขนาดของเซตการกระทำไม่เกี่ยวข้องข้อกับอัตราส่วนความไม่สมดุลเลยเนื่องจากจำนวนกฎที่อยู่ในเซตการกระทำจะขึ้นอยู่กับกระบวนการทำงานของเจเนติกอัลกอริทึม ดังนั้นจึงทำการแก้ไขเฉพาะสมการปรับค่าพารามิเตอร์ p และ \mathcal{E}

ในการปรับสมดุลอัตราการเรียนรู้จะปรับลดค่าอัตราการเรียนรู้ของกลุ่มข้อมูลที่มีข้อมูลจำนวนมากลง แต่อัตราการเรียนรู้ของกลุ่มข้อมูลที่มีข้อมูลจำนวนน้อยยังคงไว้เช่นเดิม การปรับค่าอัตราส่วนการเรียนรู้ลักษณะนี้เป็นการปรับสมดุลการเรียนรู้ของกลุ่มข้อมูลที่มีจำนวนมากเพื่อให้ใกล้เคียงกับการเรียนรู้ของกลุ่มข้อมูลที่มีจำนวนน้อย ดังตัวอย่างต่อไปนี้

สมมติให้ชุดข้อมูลมีจำนวนข้อมูลของกลุ่มที่มีจำนวนมากเป็น ir และจำนวนข้อมูลของกลุ่มข้อมูลที่มีจำนวนน้อยเป็น 1 เมื่อผ่านการเรียนรู้ $ir+1$ แล้ว XCS ก็จะได้ทำการปรับค่า p และ \mathcal{E} ของกลุ่มข้อมูลที่มีจำนวนมากเป็น ir ครั้ง ครั้งละ β/ir เมื่อครบรอบการฝึกสอนของ 1 รอบของชุดข้อมูล กลุ่มข้อมูลที่มีจำนวนมากจะมีการเรียนรู้เป็น $ir \cdot (\beta / ir) = \beta$ และ กลุ่มข้อมูลที่มีจำนวนมากจะมีการเรียนรู้เป็น β

$$p_j(t) = \begin{cases} p_j(t-1) + \psi_j(t) \times \beta \times (R - p_j(t-1)) & \text{if } a_j \text{ is major class} \\ p_j(t-1) + \beta \times (R - p_j(t-1)) & \text{if } a_j \text{ is other class} \end{cases} \quad (4.3)$$

$$\mathcal{E}_j(t) = \begin{cases} \mathcal{E}_j(t-1) + \psi_j(t) \times \beta \times (R - \mathcal{E}_j(t-1)) & \text{if } a_j \text{ is major class} \\ \mathcal{E}_j(t-1) + \beta \times (|R - p(t)| - \mathcal{E}_j(t-1)) & \text{if } a_j \text{ is other class} \end{cases} \quad (4.4)$$

เพื่อเปรียบเทียบผลลัพธ์ที่ได้จากการปรับค่าอัตราการเรียนรู้ด้วยค่า perception rate จึงนำเสนอการทดลองกับชุดข้อมูลอุปกรณ์ผสมสัญญาณขนาด 11 บิตที่ระดับความไม่สมดุลที่ระดับที่ 7 โดยกำหนดให้ XCS ที่มีการปรับค่าอัตราการเรียนรู้ด้วยค่า perception rate ให้มีเพียง 2 กฎเท่านั้นคือ กฎ ##### : 0 และกฎ ##### : 1 จากนั้นจึงเปรียบเทียบกับผลลัพธ์ที่ได้จากการทดลอง ในหัวข้อ 3.3 ได้ดังตารางต่อไปนี้

ตาราง 4.1 เปรียบเทียบค่า prediction payoff และค่า prediction error ที่ได้จากการทดลอง

Rule	$P_d(C)$	Conventional XCS		XCS with perception rate	
		p	ϵ	p	ϵ
##### : 0	128/129	1000	1.2×10^{-4}	470.90424	532.43555
##### : 1	1/129	1.2×10^{-4}	7.4×10^{-5}	580.4151	584.0643

จากตารางที่ 4.1 พบว่า XCS ที่มีการเพิ่มค่า perception rate สามารถทำให้กฎที่มีลักษณะเป็น overgeneral ไม่นิยมเพียงเข้าหากลุ่มใดกลุ่มหนึ่ง และมีค่า prediction error ที่สูงเมื่อเป็นการใช้งานทั่วไป กฎที่เป็น overgeneral จะโดนกำจัดทิ้งเนื่องจากมีค่าความผิดพลาดที่สูงมาก

4.2.3 การขยายช่วงเวลาในการทำงานของเจเนติกอัลกอริทึม

จากหัวข้อที่ 4.2.1 แสดงความสัมพันธ์ระหว่างความสัมพันธ์ระหว่างการกำหนดค่าช่วงเวลาขั้นต่ำในการทำงานของเจเนติกอัลกอริทึมต่อช่วงเวลาเฉลี่ยที่จะรับข้อมูลจากกลุ่มข้อมูลที่มีจำนวนน้อยไว้ดังนี้ การกำหนดค่าช่วงเวลาขั้นต่ำในการทำงานของเจเนติกอัลกอริทึม ควรจะกำหนดให้มีค่าไม่น้อยกว่าช่วงเวลาเฉลี่ยที่ระบบจักรกลเรียนรู้จะรับข้อมูลจากกลุ่มข้อมูลที่มีจำนวนน้อย

ช่วงเวลาเฉลี่ยที่ระบบจักรกลเรียนรู้จะรับข้อมูลจากกลุ่มข้อมูลที่มีจำนวนน้อยสามารถคำนวณได้จากอัตราส่วนความไม่สมดุลของข้อมูล จำนวนครั้งที่รับข้อมูลจากกลุ่มข้อมูลที่มีจำนวนมาก ต่อการรับข้อมูลจากกลุ่มข้อมูลที่มีจำนวนน้อย 1 ครั้ง แต่อัตราส่วนความไม่สมดุลของข้อมูลคำนวณจากจำนวนข้อมูลจากกลุ่มข้อมูลที่มีจำนวนน้อย ต่อจำนวนข้อมูลจากกลุ่มข้อมูลที่มีจำนวนมาก ดังนั้นช่วงเวลาเฉลี่ยที่ระบบจักรกลเรียนรู้จะรับข้อมูลจากกลุ่มข้อมูลที่มีจำนวนน้อยจึงมีค่าเป็นส่วนกลับของอัตราส่วนความไม่สมดุล แต่ Butz และ Wilson แนะนำแนวทางในการกำหนดค่าช่วงเวลาขั้นต่ำในการทำงานของเจเนติกอัลกอริทึมในบทความวิจัยเรื่อง An Algorithmic Description of XCS [14] ว่าการกำหนดค่าช่วงเวลาขั้นต่ำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ต่ำในการทำงานของเจเนติกอัลกอริทึมไม่ควรต่ำกว่า 25 ดังนั้นจึงสามารถประมาณค่าช่วงเวลานั้นต่ำในการทำงานของเจเนติกอัลกอริทึมได้ ดังสมการต่อไปนี้

$$new_ \theta_{GA} = \begin{cases} 25 & \text{if } ir \leq 25 \\ ir & \text{otherwise} \end{cases} \quad (4.15)$$

โดยที่ $new_ \theta_{GA}$ คือค่าช่วงเวลานั้นต่ำในการทำงานของเจเนติกอัลกอริทึมที่คำนวณได้ใหม่

เพื่อให้กระบวนการจำแนกมีประสิทธิภาพสูงที่สุด ควรกำหนดการทำงานของเจเนติกอัลกอริทึมเกิดขึ้นบนทุกๆ กฎที่อยู่ในระบบอย่างเท่าเทียมกัน โดยการกำหนดค่าช่วงเวลานั้นต่ำในการทำงานของเจเนติกอัลกอริทึมควรมีค่าเดียวกันทั้งระบบ แต่เนื่องจากค่าช่วงเวลานั้นต่ำในการทำงานของเจเนติกอัลกอริทึมที่นำเสนอมีการปรับค่าตามอัตราส่วนความไม่สมดุลของกลุ่มข้อมูล ดังนั้นจึงใช้ Widrow-Hoff delta rule ในการปรับค่าช่วงเวลานั้นต่ำในการทำงานของเจเนติกอัลกอริทึม ดังสมการต่อไปนี้

$$\theta_{GA}(t) = \theta_{GA}(t-1) + \beta \times (new_ \theta_{GA} - \theta_{GA}(t-1)) \quad (4.16)$$

โดยที่ $\theta_{GA}(t)$ คือค่าช่วงเวลานั้นต่ำในการทำงานของเจเนติกอัลกอริทึมในช่วงเวลาปัจจุบัน
 $\theta_{GA}(t-1)$ คือค่าช่วงเวลานั้นต่ำในการทำงานของเจเนติกอัลกอริทึมในช่วงเวลาก่อนหน้า
 β คืออัตราเรียนรู้สำหรับปรับค่าช่วงเวลานั้นต่ำในการทำงานของเจเนติกอัลกอริทึม

4.3 ชุดข้อมูล (Datasets)

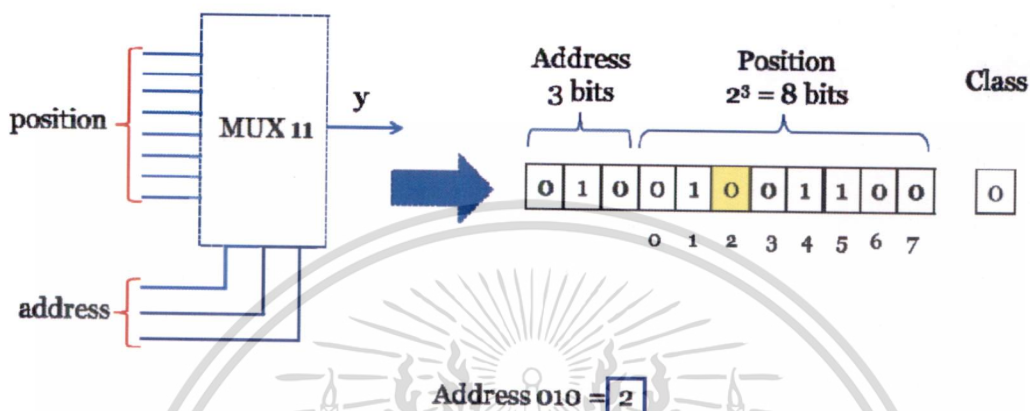
ในหัวข้อนี้นำเสนอ รายละเอียดของชุดข้อมูลที่ใช้ในการทดลอง โดยแบ่งเป็น ชุดข้อมูลสังเคราะห์ (Simulate Dataset) และ ชุดข้อมูลจริง (Real Dataset) โดยชุดข้อมูลสังเคราะห์ เป็นชุดข้อมูลที่มีความนิยมในการทดสอบ ระบบจักรกลเรียนรู้

4.3.1 ชุดข้อมูลสังเคราะห์ (Simulate Datasets)

ชุดข้อมูลสังเคราะห์ที่ใช้ในการทดลองนี้มีชื่อว่า ชุดข้อมูลปัญหา มัลติเพล็กซ์เซอร์ (Multiplexer Problem) นำเสนอโดย Wilson [2] ชุดข้อมูลปัญหา มัลติเพล็กซ์เซอร์นี้สังเคราะห์ขึ้นจากการจำลองการทำงานของอุปกรณ์ผสมสัญญาณ (Multiplexer) ให้อยู่ในรูปแบบของสายข้อมูลไบนารี (Binary Stream) โดยมีส่วนประกอบ 2 ส่วนคือ ตัวชี้ตำแหน่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนำไปเผยแพร่โดยไม่ได้รับอนุญาตเป็นการฝ่าฝืน
 ไม่ว่าการณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Address) เป็นส่วนต้นของสายข้อมูล มีขนาด l บิต และส่วนข้อมูล (Data) เป็นส่วนท้ายของข้อมูล มีขนาด 2^l บิต เมื่อรวมทั้งสองส่วนเข้าด้วยกันแล้วข้อมูลของมัลติเพล็กซ์เซอร์จึงมีขนาด $l + 2^l$ บิต และสามารถกำหนดกลุ่มของข้อมูลมัลติเพล็กซ์เซอร์ได้จากข้อมูลในตำแหน่งที่ตัวชี้อ้างอิง ดังรูป 5.1



รูป 4.1 ตัวอย่างแสดงการจำลองการทำงานของอุปกรณ์ผสมสัญญาณขนาด 11 บิต เป็นข้อมูลปัญหา มัลติเพล็กซ์เซอร์ 11 บิต

ตาราง 4.2 ตัวอย่างชุดข้อมูลปัญหา Multiplexer 11 บิต

Class 0		Class 1	
000 00000000	001 00000000	000 10000000	001 01000000
010 00000000	011 00000000	010 00100000	011 00010000
100 00000000	101 00000000	100 00001000	101 00000100
110 00000000	111 00000000	110 00000010	111 00000001

4.3.2 ชุดข้อมูลจริง (Real Dataset)

ชุดข้อมูลจริงที่ใช้ในการทดลองในวิทยานิพนธ์ฉบับนี้เป็นบางส่วนของชุดข้อมูลมาตรฐานที่ใช้ในการทดสอบประสิทธิภาพในการจำแนกประเภทที่มีชื่อว่า UCI datasets โดยสามารถดาวน์โหลดได้จาก <http://archive.ics.uci.edu/ml/> ยกเว้นชุดข้อมูล Tao Grid [13] เป็นชุดข้อมูลที่ได้จากการสังเคราะห์ขึ้นมาจากรูปสัญลักษณ์เต๋า โดยชุดข้อมูลที่ใช้ในการทดสอบนี้เป็นชุดข้อมูลที่นิยมใช้งานกันอย่างกว้างขวาง รายละเอียดของชุดข้อมูลที่นำมาใช้ในการทดสอบมีดังตารางต่อไปนี้

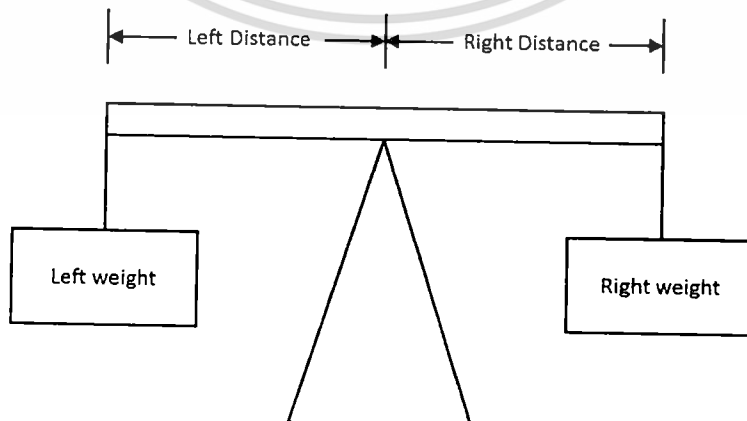
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 4.3 ชุดข้อมูลที่นำมาใช้ในการทดลองเพื่อวัดค่าประสิทธิภาพ

Dataset Name	#Instances	Attributes				#Classes
		binary	nominal	ordinal	real	
Balance-Scale	625	0	4	0	0	3
Breast Cancer Wisconsin	699	0	0	9	0	2
Diabetes	768	0	0	6	2	2
Heart-Statlog	270	3	3	1	6	2
Glass	214	0	0	0	9	7
Wine	178	0	0	0	13	3
Tao Grid	1888	0	0	2	0	2

ในชุดข้อมูลจริงที่ใช้ในการทดสอบนี้เป็นชุดข้อมูล มีชุดข้อมูล Balance-Scale และชุดข้อมูล Tao Grid ที่เป็นชุดข้อมูลสังเคราะห์ มีคุณลักษณะพิเศษดังนี้

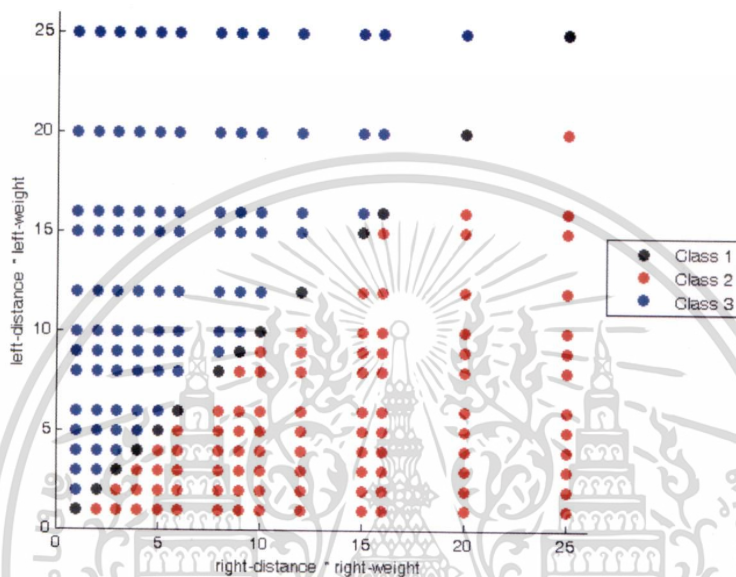
ชุดข้อมูล Balance-Scale เป็นชุดข้อมูลที่จำลองการทำงานของตราซั่ง 2 แขน ดังรูป 4.2 ชุดข้อมูล Balance-Scale ประกอบด้วยพารามิเตอร์ 4 ตัวได้แก่ น้ำหนักถ่วงทางซ้าย (left weight) ความยาวของแขนซ้าย (left distance) น้ำหนักถ่วงทางขวา (right weight) และความยาวของแขนขวา (right distance) โดยมีหลักการในการจำแนกกลุ่มข้อมูลจาก ผลคูณของน้ำหนักถ่วงทางซ้ายและความยาวของแขนซ้าย เทียบกับ ผลคูณของน้ำหนักถ่วงทางขวาและความยาวของแขนขวา โดยถ้าผลคูณของฝั่งซ้าย มากกว่า ผลคูณของฝั่งขวา จะจัดว่าเป็นกลุ่มข้อมูลเอียงซ้าย (Left : L) และถ้าผลคูณของฝั่งซ้าย น้อยกว่า ผลคูณของฝั่งขวา จะจัดว่าเป็นกลุ่มข้อมูลเอียงขวา (Right : R) และ ถ้าผลคูณของฝั่งซ้าย เท่ากับ ผลคูณของฝั่งขวา จะจัดเป็นกลุ่มข้อมูล สมดุล (Balance : B)



รูป 4.2 แบบจำลองตราซั่งสองแขน

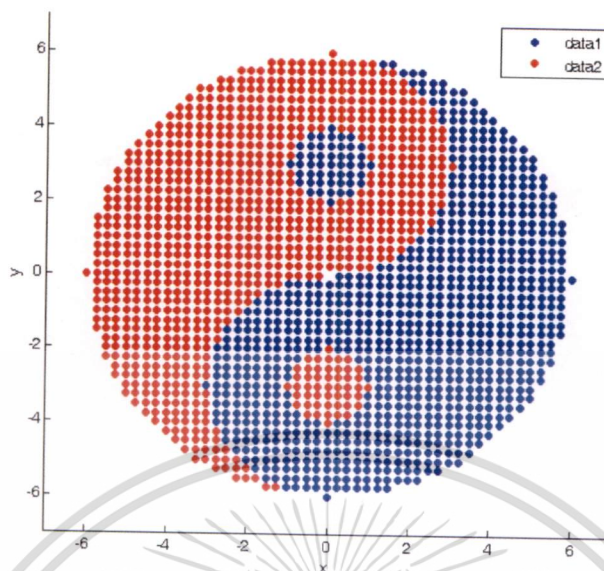
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณลักษณะพิเศษของชุดข้อมูล Balance-Scale อยู่ที่การกระจายตัวของชุดข้อมูล โดยลักษณะการกระจายตัวของกลุ่มข้อมูล B จะมีลักษณะเป็นเส้นตรง ดังรูป 4.3 และกลุ่มข้อมูล B มีจำนวนข้อมูลน้อยมากเมื่อเทียบกับจำนวนข้อมูลของกลุ่มข้อมูล L หรือ R โดยมีอัตราส่วนเป็น 1 ต่อ 5.88 เท่า



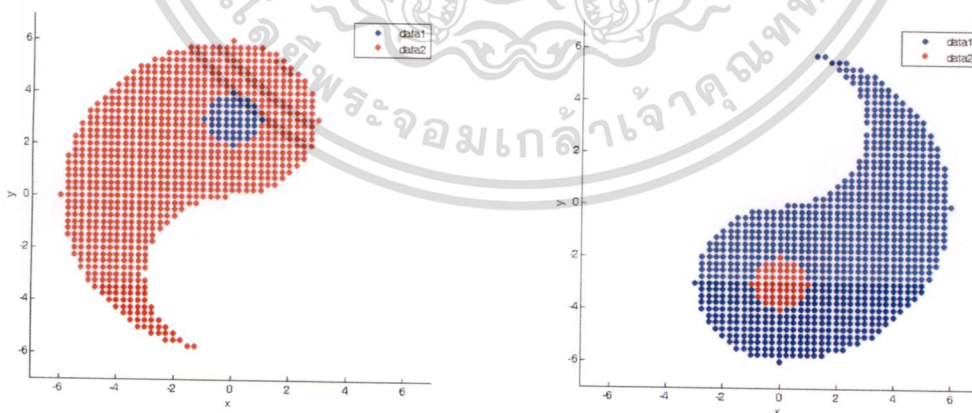
รูป 4.3 กราฟการกระจายตัวของชุดข้อมูล Balance-Scale

ชุดข้อมูล Tao Grid เป็นชุดข้อมูลที่ตั้งคราะห์ขึ้นให้มีลักษณะการกระจายตัวเหมือนสัญลักษณ์ของลัทธิเต๋า ดังรูป 4.4 ชุดข้อมูล Tao Grid ประกอบด้วย 2 กลุ่มข้อมูล ขาวและดำ แทนด้วยในรูป 4.5 แทนด้วยสีแดงและสีดำ



รูป 4.4 กราฟการกระจายตัวของชุดข้อมูล Tao Grid

คุณลักษณะพิเศษของชุดข้อมูล Tao คือ ถึงแม้ว่าชุดข้อมูล Tao Grid จะเป็นชุดข้อมูลอัตราความไม่สมดุลเป็น 1 แต่ว่า ถ้าแบ่งแยกพิจารณาเฉพาะกลุ่มข้อมูลสีขาวที่ปนอยู่ในสีดำเทียบกับกลุ่มข้อมูลสีดำโดยรวม ดังรูป 4.4 ก. หรือกลุ่มข้อมูลสีดำที่ปนอยู่ในกลุ่มข้อมูลสีขาวโดยรวม จะพบว่าม้อตราส่วนความไม่สมดุลสูงถึง 15.276 คุณลักษณะนี้เรียกว่า ความไม่สมดุลในพื้นที่ย่อย (imbalance in subspace)



ก. กราฟการกระจายตัวของกลุ่มข้อมูลสีดำที่ปนอยู่ในสีแดง

ข. กราฟการกระจายตัวของกลุ่มข้อมูลสีแดงที่ปนอยู่ในสีดำ

รูป 4.5 กราฟการกระจายตัวของชุดข้อมูล Tao Grid ในส่วนของพื้นที่ย่อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อวัดประสิทธิภาพการจำแนกกลุ่มข้อมูลที่ระดับความไม่สมดุลที่แตกต่างกัน จึงทำการแบ่งชุดข้อมูลออกเป็น 2 ลักษณะได้แก่ ชุดข้อมูลที่มีกลุ่มข้อมูลตั้งแต่ 3 กลุ่มขึ้นไป (Multiple Class Datasets) ชุดข้อมูลที่มีกลุ่มข้อมูล 2 กลุ่ม (Binary Class Datasets) โดยมีรายละเอียดดังต่อไปนี้

ตาราง 4.4 การกระจายตัวของกลุ่มข้อมูลในชุดข้อมูลที่มีตั้งแต่ 3 กลุ่มขึ้นไป

Dataset Name	#Instances	#Classes	Class distribution $\#C_1 / \#C_2 / \dots / \#C_n$
Balance-Scale	625	3	49 / 288 / 288
Glass	214	7	70 / 76 / 17 / 0 / 13 / 9 / 29
Wine	178	3	59 / 71 / 48

ตาราง 4.4 แสดงการกระจายตัวของกลุ่มข้อมูลในชุดข้อมูลที่มีกลุ่มข้อมูลตั้งแต่ 3 กลุ่มขึ้นไป ตารางนี้ประกอบด้วย ชื่อของกลุ่มข้อมูล จำนวนข้อมูลในชุดข้อมูล จำนวนกลุ่มข้อมูล และการกระจายตัวของข้อมูลในแต่ละกลุ่มข้อมูล เช่น ชุดข้อมูล Balance-Scale มีข้อมูลทั้งหมด 625 ตัว แบ่งเป็น 3 กลุ่ม โดยกลุ่มที่ 1 มีข้อมูล 49 ตัว กลุ่มที่ 2 มีข้อมูล 288 ตัว และกลุ่มที่ 3 มีข้อมูล 288 ตัว

เนื่องจากชุดข้อมูลที่มี 2 กลุ่มมีจำนวนน้อยจึงได้แปลงชุดข้อมูลที่มีมากกว่า 2 กลุ่มให้เป็นชุดข้อมูลที่มี 2 กลุ่มหลายๆ ชุดข้อมูล โดยกำหนดให้กลุ่มข้อมูลที่สนใจเห็นกลุ่มข้อมูลที่มีจำนวนน้อย และกลุ่มข้อมูลที่เหลือรวมเป็นกลุ่มเดียวกันเป็นกลุ่มข้อมูลที่มีจำนวนมาก ดังตาราง 4.4 โดยมีตัวอย่างในการแปลงชุดข้อมูลดังต่อไปนี้

ตัวอย่างการแปลงชุดข้อมูลเป็น Wine ซึ่งเป็นชุดข้อมูลที่มี 3 กลุ่มข้อมูล ให้เป็นชุดข้อมูลที่มี 2 กลุ่มข้อมูล จำนวน 3 ชุดข้อมูล ชุดข้อมูล Wine ประกอบด้วยกลุ่มข้อมูล 3 กลุ่ม โดยกลุ่มที่ 1 มีข้อมูล 59 ตัว กลุ่มที่ 2 มีข้อมูล 71 ตัว และกลุ่มที่ 3 มีข้อมูล 48 ตัว สามารถแปลงเป็นชุดข้อมูลที่มี 2 กลุ่มข้อมูลได้ 3 ชุด ได้ดังนี้

- ชุดข้อมูลที่มีกลุ่มข้อมูลที่ 1 จากชุดข้อมูล Wine เป็นกลุ่มข้อมูลที่มีจำนวนน้อย เรียกชุดข้อมูลนี้ว่า “Wine disc 1” ชุดข้อมูลนี้จะมีจำนวนข้อมูลในกลุ่มที่มีจำนวนน้อย 59 ตัว และรวมกลุ่มข้อมูลที่ 2 และ 3 เป็นกลุ่มข้อมูลที่มีจำนวนมาก มีข้อมูลเป็น $71+48 = 119$ ข้อมูล
- ชุดข้อมูลที่มีกลุ่มข้อมูลที่ 2 จากชุดข้อมูล Wine เป็นกลุ่มข้อมูลที่มีจำนวนน้อย เรียกชุดข้อมูลนี้ว่า “Wine disc 2” ชุดข้อมูลนี้จะมีจำนวนข้อมูลในกลุ่มที่มีจำนวน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

น้อยเป็น 71 ข้อมูล และรวมกลุ่มข้อมูลที่ 1 และ 3 เป็นกลุ่มข้อมูลที่มีจำนวนมาก เป็น $59+48 = 107$ ข้อมูล

- ชุดข้อมูลที่มีกลุ่มข้อมูลที่ 3 จากชุดข้อมูล Wine เป็นกลุ่มข้อมูลที่มีจำนวนน้อย เรียกชุดข้อมูลนี้ว่า “Wine disc 3” ชุดข้อมูลนี้จะมีจำนวนข้อมูลในกลุ่มที่มีจำนวนน้อย 48 ตัว และรวมกลุ่มข้อมูลที่ 1 และ 2 เป็นกลุ่มข้อมูลที่มีจำนวนมาก มีข้อมูล $59+71 = 130$ ตัว

ตาราง 4.5 รายละเอียดของชุดข้อมูลที่ใช้ที่มี 2 กลุ่มข้อมูล

Dataset	#Ins.	#Major	#Minor	<i>Ir</i>
Breast Cancer Wisconsin	699	458	241	1.90 : 1
Tao Grid	1888	944	944	1.00 : 1
Diabetes	768	500	268	1.87 : 1
Heart-Statlog	270	150	120	1.25 : 1
Balance-Scale dist1	625	576	49	11.75 : 1
Balance-Scale dist2	625	337	288	1.17 : 1
Balance-Scale dist3	625	337	288	1.17 : 1
Glass dist1	214	144	70	2.06 : 1
Glass dist2	214	138	76	1.82 : 1
Glass dist3	214	197	17	11.59 : 1
Glass dist7	214	185	29	6.38 : 1
Wine dist1	178	119	59	2.02 : 1
Wine dist2	178	107	71	1.50 : 1
Wine dist3	178	130	48	2.70 : 1

สำหรับชุดข้อมูล Glass เป็นชุดข้อมูลที่มีกลุ่มข้อมูลทั้งหมด 7 กลุ่ม แต่มีชุดอยู่ 3 กลุ่ม ข้อมูลที่มีจำนวนน้อยมากๆ ได้แก่ กลุ่มข้อมูลที่ 4 ซึ่งมีจำนวนข้อมูลในกลุ่มเป็น 0 กลุ่มข้อมูลที่ 5 ซึ่งมีจำนวนข้อมูลในกลุ่มเป็น 13 และกลุ่มข้อมูลที่ 6 ซึ่งมีจำนวนข้อมูลในกลุ่มเป็น 9 กลุ่มข้อมูลทั้ง 3 นี้เป็นกลุ่มข้อมูลที่มีจำนวนน้อยจนทำให้ไม่สามารถวัดประสิทธิภาพด้วย 10 fold-cross validation technique ได้อย่างแม่นยำ ดังนั้นจึงเลือกใช้เฉพาะ กลุ่มข้อมูลที่ 1 2 3 และ 7 มาใช้เป็นฐานในการสร้างชุดข้อมูลชุดใหม่เพื่อใช้ในการฝึกสอนและทดสอบด้วยวิธี 10 fold-cross validation technique

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 การวัดประสิทธิภาพ

ในการวัดประสิทธิภาพของจักรกลเรียนรู้ทั่วไปจะใช้การวัดค่าความถูกต้องมาตรฐาน (Standard accuracy) โดยวัดค่าความถูกต้องโดยรวมของระบบ (total accuracy) จากอัตราส่วนของจำนวนครั้งที่จักรกลเรียนรู้จำแนกกลุ่มข้อมูล ได้ถูกต้องจำนวนของข้อมูลทั้งหมด เนื่องจากชุดข้อมูลที่ไม่มีสมดุลมีจำนวนข้อมูลในบางกลุ่มเป็นจำนวนมากทำให้ผลลัพธ์ที่ได้จากการวัดค่าความถูกต้องมาตรฐานจะขึ้นอยู่กับค่าความถูกต้องของกลุ่มที่มีข้อมูลจำนวนมาก ดังตัวอย่างต่อไปนี้

ตัวอย่างเช่น สมมติมีชุดข้อมูล 1 ชุด ประกอบไปด้วย 2 กลุ่มข้อมูล ได้แก่ กลุ่มข้อมูล “ก” และกลุ่มข้อมูล “ข” กลุ่มข้อมูล “ก” มีจำนวนข้อมูลในกลุ่มคิดเป็นร้อยละ 90 ของข้อมูลทั้งหมด และกลุ่มข้อมูล “ข” มีจำนวนข้อมูลในกลุ่มคิดเป็นร้อยละ 10 ของข้อมูลทั้งหมด ถ้าจักรกลเรียนรู้สามารถจำแนกกลุ่มข้อมูลของข้อมูลที่ระบุกลุ่มข้อมูลเป็น “ก” ได้อย่างถูกต้องทั้งหมด แต่ไม่สามารถจำแนกกลุ่มข้อมูลของข้อมูลที่ระบุกลุ่มข้อมูลเป็น “ข” ได้เลย ค่าความถูกต้องรวมที่ได้จากการจำแนกของจักรกลเรียนรู้มีค่าเป็นร้อยละ 90 แต่ถ้าจักรกลเรียนรู้สามารถจำแนกกลุ่มข้อมูลของข้อมูลที่ระบุกลุ่มข้อมูลเป็น “ข” ได้อย่างถูกต้องทั้งหมด ในขณะที่ไม่สามารถจำแนกกลุ่มข้อมูลของข้อมูลที่ระบุกลุ่มข้อมูลเป็น “ก” ได้เลย ค่าความถูกต้องรวมที่ได้จากการจำแนกของจักรกลเรียนรู้มีค่าเป็นร้อยละ 10

จากตัวอย่างข้างต้นแสดงให้เห็นว่าค่าประสิทธิภาพที่ได้จากการวัดค่าความถูกต้องมาตรฐานจะคล้อยตามค่าความถูกต้องของกลุ่มข้อมูลที่มีจำนวนข้อมูลมากที่สุดเป็นส่วนมาก ดังนั้นการวัดค่าความถูกต้องมาตรฐานจึงไม่เหมาะสมที่จะใช้ในการวัดประสิทธิภาพของจักรกลเรียนรู้ที่ใช้งานกับชุดข้อมูลที่ไม่มีสมดุล ในการวัดค่าประสิทธิภาพของจักรกลเรียนรู้ที่ใช้งานกับชุดข้อมูลที่ไม่มีสมดุลมีวิธีการที่นิยมใช้กันคือ ค่าเฉลี่ยเรขาคณิต (geometric mean) [14,15]

4.4.1 ค่าเฉลี่ยเรขาคณิต (Geometric mean)

ค่าเฉลี่ยเรขาคณิตเป็นวิธีวัดประสิทธิภาพของจักรกลเรียนรู้ที่นิยมใช้กันมากในการวัดค่าประสิทธิภาพของจักรกลเรียนรู้กับชุดข้อมูลที่มีความไม่สมดุล ค่าเฉลี่ยเรขาคณิตเป็นค่ากลาง (mean) หรือค่าเฉลี่ย (average) รูปแบบหนึ่งซึ่งได้จากการหารากที่ n ของผลคูณของข้อมูลที่มีค่ามากกว่า 0 จำนวน n ตัว ดังสมการต่อไปนี้

$$\left(\prod_{i=1}^n a_i\right)^{1/n} = \sqrt[n]{a_1 \cdot a_2 \cdots a_n} \quad (5.1)$$

การวัดค่าประสิทธิภาพของจักรกลเรียนรู้กับชุดข้อมูลที่ไม่สมดุลจะแยกคำนวณหาค่าความถูกต้องในการจำแนกของแต่ละกลุ่มข้อมูลก่อน แล้วจึงนำค่าความถูกต้องในการจำแนกของแต่ละกลุ่มมาคำนวณหาค่าเฉลี่ยเรขาคณิตเพื่อแสดงค่าประสิทธิภาพโดยรวมของการจำแนกทั้งหมด

การคำนวณหาค่าความถูกต้องของแต่ละกลุ่มข้อมูลสามารถคำนวณจากจำนวนครั้งที่จักรกลเรียนรู้จำแนกข้อมูลในกลุ่มข้อมูลนั้นๆ ถูกต้อง ต่อจำนวนข้อมูลทั้งหมดในกลุ่มข้อมูลที่พิจารณา ดังตัวอย่างต่อไปนี้

ตัวอย่าง สมมุติให้ชุดข้อมูลประกอบไปด้วยกลุ่มข้อมูล 2 กลุ่ม ได้แก่กลุ่มข้อมูล “+” และ กลุ่มข้อมูล “-” โดยสามารถแบ่งกลุ่มของผลลัพธ์ที่ได้จากการจำแนกออกเป็น 4 กรณี ดังตาราง 4.6

ตาราง 4.6 การประเมินผลการทำนาย (Confusion matrix)

		กลุ่มข้อมูลที่จำแนก	
		+	-
กลุ่มข้อมูลที่ระบุ	+	a	b
	-	c	d

จากตาราง 4.6 สามารถคำนวณดัชนีชี้วัดประสิทธิภาพ ได้ 4 ชนิดดังต่อไปนี้

1. True Positive Rate คือ อัตราความถูกต้องในการจำแนกข้อมูลจากกลุ่มข้อมูล “+” ได้ถูกต้องต่อจำนวนข้อมูลที่ระบุกลุ่มข้อมูลเป็น “+” ทั้งหมด เทียบกับในตาราง 5.4 ค่า True Positive Rate จะมีค่าเป็น $a / (a + b)$
2. False Positive Rate คือ อัตราความถูกต้องในการจำแนกข้อมูลจากกลุ่มข้อมูล “+” ผิด ต่อจำนวนข้อมูลที่ระบุกลุ่มข้อมูลเป็น “+” ทั้งหมด เทียบกับในตาราง 5.4 ค่า False Positive Rate จะมีค่าเป็น $b / (a + b)$
3. True Negative Rate คือ อัตราความถูกต้องในการจำแนกข้อมูลจากกลุ่มข้อมูล “-” ได้ถูกต้องต่อจำนวนข้อมูลที่ระบุกลุ่มข้อมูลเป็น “-” ทั้งหมด เทียบกับในตาราง 5.4 ค่า False Positive Rate จะมีค่าเป็น $c / (c + d)$
4. False Negative Rate คือ อัตราความถูกต้องในการจำแนกข้อมูลจากกลุ่มข้อมูล “-” ผิด ต่อจำนวนข้อมูลที่ระบุกลุ่มข้อมูลเป็น “-” ทั้งหมด เทียบกับในตาราง 5.4 ค่า False Positive Rate จะมีค่าเป็น $d / (c + d)$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.2 k-Fold Cross-Validation technique

k-Fold Cross-Validation เป็นวิธีแบ่งชุดข้อมูลออกเป็นชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบที่ได้รับความนิยมเป็นอย่างมาก[15] เนื่องจากการใช้งานจากจักรกลเรียนรู้ต้องมีการแบ่งชุดข้อมูลออกเป็นชุดข้อมูลฝึกสอน (Training dataset) และชุดข้อมูลทดสอบ (Testing dataset) แต่ในบางครั้งการเลือกชุดข้อมูลก็ส่งผลกระทบต่อประสิทธิภาพในการจำแนกข้อมูล โดยถ้าเลือกชุดข้อมูลที่ใช้เป็นชุดข้อมูลทดสอบเป็นชุดข้อมูลที่แบ่งไม่ดี เช่น ไม่มีสัญญาณรบกวน (noise) เลย หรือไม่กระจายครอบคลุมทั้งชุดข้อมูล ก็จะส่งผลกระทบต่อค่าประสิทธิภาพในการจำแนกข้อมูลของจักรกลเรียนรู้ที่วัดได้ ดังนั้นจึงมีการคิด k-fold cross validation ขึ้นมาแก้ปัญหานี้

k-Fold Cross validation มีการทำงานโดยการสร้างคู่ของชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ k คู่ จากชุดข้อมูล 1 ชุด โดยแบ่งข้อมูลออกเป็น k กลุ่ม (k-Fold) และทำการกำหนดให้ k-1 กลุ่มเป็นชุดข้อมูลฝึกสอน และอีก 1 กลุ่มเป็นชุดข้อมูลทดสอบ คู่ของชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบชุดที่ 1 จะกำหนดให้ข้อมูลกลุ่มที่ 1 เป็นข้อมูลชุดทดสอบ และข้อมูลกลุ่มที่เหลือ k-1 กลุ่มเป็นชุดข้อมูลฝึกสอน และคู่ของชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบชุดแรกชุดที่ 2 จะกำหนดให้ข้อมูลกลุ่มที่ 2 เป็นข้อมูลชุดทดสอบ และข้อมูลกลุ่มที่เหลือเป็นชุดข้อมูลฝึกสอน สลับอย่างนี้ไปเรื่อยๆ จนถึงคู่ของชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบชุดที่ k วิธีการนี้จะทำให้ได้คู่ของชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ k คู่ จากนั้นจะต้องทำการฝึกสอนและทดสอบจักรกลเรียนรู้ k ครั้งแล้วทำการหาค่าเฉลี่ยของค่าความถูกต้องในแต่ละครั้ง

4.5 การทดลอง

ในหัวข้อนี้จะอธิบายการทดลองเพื่อวัดค่าประสิทธิภาพของ XCS โดยแบ่งการทดลองออกเป็น 3 ส่วนหลักๆ ได้แก่ การทดลองกับชุดสังเคราะห์ และการทดลองกับชุดข้อมูลจริง 2 กลุ่ม การทดลองกับชุดข้อมูลจริงตั้งแต่ 3 กลุ่มขึ้นไป โดยมีรายละเอียดดังต่อไปนี้

4.5.1 การทดลองกับชุดข้อมูลสังเคราะห์

การทดลองกับชุดข้อมูลสังเคราะห์ เป็นการทดลองเพื่อศึกษาผลกระทบของประสิทธิภาพในการจำแนกกลุ่มข้อมูลของ XCS ที่มีตัวแปรที่ปรับค่าได้ กับชุดข้อมูลที่มีระดับความไม่สมดุลแตกต่างกัน 11 ระดับ โดยทำการทดลองเพื่อศึกษาความสำคัญในการปรับค่าของตัวแปรแต่ละตัว ซึ่งแบ่งเป็นการทดลองย่อยๆ 4 การทดลอง ดังต่อไปนี้

1) ทดลองด้วย XCS มาตรฐาน (Conventional XCS)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 2) ทดลองด้วย XCS ที่มีการปรับค่า GA threshold (XCS with adaptive GA threshold)
- 3) ทดลองด้วย XCS ที่มีการปรับค่าอัตราการรับรู้ (XCS with adaptive perception rate)
- 4) ทดลองด้วย XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold (XCS with adaptive perception rate and GA threshold)

ในการทดลองแต่ละครั้งมีการกำหนดค่าสถานะในการทดลองเป็นดังต่อไปนี้ กำหนดจำนวนหน่วยข้อมูลที่ใช้ในการฝึกสอนเป็น 5,000,000 หน่วยข้อมูล จำนวนประชากรของกฎมีค่าเป็น 2000 ค่าอัตราการเรียนรู้เริ่มต้นมีค่าเป็น 0.2 ค่า GA Threshold เริ่มต้นเป็น 25 ค่า crossing over threshold เป็น 0.8 ค่า mutation threshold เป็น 0.04 และกำหนดค่าตอบแทนสำหรับการจำแนกกลุ่มข้อมูล เป็น 2 ค่า ได้แก่ ค่าตอบแทนสูงสุด R_{Max} เป็น 1000 สำหรับกรณีที่ระบบสามารถจำแนกกลุ่มข้อมูลได้ถูกต้อง และค่าตอบแทนต่ำสุด R_{Min} เป็น 0 สำหรับกรณีที่ระบบจำแนกกลุ่มข้อมูลผิดพลาด ตามงานวิจัยของ [11]

ในส่วนนี้จะทดลองกับชุดข้อมูลปัญหาหมัดติเพ็ล็กซ์เซอร์ที่ระดับความไม่สมดุลต่างๆ กัน 11 ระดับ ตั้งแต่ระดับความไม่สมดุลที่ $i = 0$ ซึ่งเป็นระดับที่สมดุล ไปจนถึงระดับความไม่สมดุลที่ $i = 10$ ซึ่งมีอัตราส่วนระหว่าง จำนวนหน่วยข้อมูลในกลุ่มข้อมูลที่มีจำนวนมาก ต่อจำนวนหน่วยข้อมูลในกลุ่มข้อมูลที่มีจำนวนน้อย เป็น $2^{10} = 1024$ เท่า

ตาราง 4.7 แสดงค่าความถูกต้องในการจำแนกกลุ่มข้อมูล '1' และส่วนเบี่ยงเบนมาตรฐานของ XCS 3 ชนิดโดยเฉลี่ยจากการทดลองกับชุดข้อมูลปัญหาหมัดติเพ็ล็กซ์เซอร์ 11 บิตด้วยวิธี 10-Flod Cross-validation (\uparrow หมายถึงประสิทธิภาพในการจำแนกของจักรกลเรียนรู้ดีขึ้นอย่างมีนัยสำคัญที่ดัชนีความเชื่อมั่น 0.05 เมื่อเปรียบเทียบกับ XCS มาตรฐาน และ \downarrow หมายถึงประสิทธิภาพในการจำแนกจักรกลเรียนรู้ด้อยลงอย่างมีนัยสำคัญที่ดัชนีความเชื่อมั่น 0.05 เมื่อเปรียบเทียบกับ XCS)

จากตาราง 4.7 จะเห็นว่าค่าประสิทธิภาพในการจำแนกกลุ่มข้อมูลที่ไม่สมดุลของ XCS ที่มีการปรับค่า GA threshold ไม่สามารถเพิ่มประสิทธิภาพในการจำแนกกลุ่มข้อมูลที่ไม่สมดุลให้ดีขึ้นกว่า XCS มาตรฐาน

ตาราง 4.7 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูล '1' และส่วนเบี่ยงเบนมาตรฐานของ XCS

Imbalance level	Accuracy of classification group label '1'			
	Conventional XCS	XCS with adaptive GA	XCS with adaptive perception rate	XCS with adaptive perception rate and GA
$i = 0$	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
$i = 1$	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
$i = 2$	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
$i = 3$	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
$i = 4$	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
$i = 5$	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
$i = 6$	0.80 ± 0.20	0.80 ± 0.20	0.90 ± 0.00	$1.00 \pm 0.00 \uparrow$
$i = 7$	0.00 ± 0.00	0.00 ± 0.00	$0.80 \pm 0.00 \uparrow$	$1.00 \pm 0.00 \uparrow$
$i = 8$	0.00 ± 0.00	0.00 ± 0.00	0.05 ± 0.15	$1.00 \pm 0.00 \uparrow \uparrow$
$i = 9$	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	$1.00 \pm 0.00 \uparrow \uparrow$
$i = 10$	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	$1.00 \pm 0.00 \uparrow \uparrow$

ในขณะที่ XCS ที่มีการปรับค่าอัตราการเรียนรู้ สามารถเพิ่มประสิทธิภาพในการจำแนกกลุ่มข้อมูลที่ไม่สมดุลเพิ่มมากขึ้น เมื่ออัตราส่วนความไม่สมดุลเพิ่มมากขึ้นจนถึงระดับที่ $i=7$ เนื่องจากการปรับค่าอัตราการเรียนรู้จะทำให้การคำนวณค่า prediction reward, prediction error และค่า fitness ที่สมดุลระหว่างกันระหว่างกลุ่มข้อมูล 1 ที่มีจำนวนน้อย กับ กลุ่มข้อมูล 0 ที่มีจำนวนมาก ได้ดีขึ้น

XCS ที่มีการปรับค่าอัตราการเรียนรู้ และ GA threshold สามารถจำแนกกลุ่มข้อมูลที่ไม่สมดุลได้ทุกๆ ระดับความไม่สมดุลที่ทำการทดลอง เนื่องจากการปรับค่า เนื่องจากการปรับค่าอัตราการเรียนรู้จะทำให้การคำนวณค่า prediction reward, prediction error และค่า fitness ให้เหมาะสมกับความไม่สมดุล แต่การปรับค่าอัตราการเรียนรู้จำเป็นต้องใช้เวลาในการปรับค่าให้เหมาะสม ดังนั้นค่า GA threshold ตามอัตราส่วนความไม่สมดุลซึ่งเป็นการปรับให้เข้ากับระยะเวลาในการเรียนรู้ของค่าอัตราการเรียนรู้ จึงทำให้ XCS สามารถจำแนกกลุ่มข้อมูลที่ไม่สมดุลได้ดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.2 การทดลองกับชุดข้อมูลจริง 2 กลุ่ม

การทดลองกับชุดข้อมูลจริงเป็นการทดลองเพื่อวัดค่าประสิทธิภาพการจำแนกกลุ่มข้อมูลในสถานะการใช้งานจริง เนื่องจากการใช้งานจริงมีปัจจัยหลายอย่างที่ทำให้เกิดการเรียนรู้ผิดพลาดได้ เช่น สัญญาณรบกวน Noise กลุ่มข้อมูลมีการคาบเกี่ยวกัน (overlapping) ฯลฯ โดยชุดข้อมูลที่ใช้ในการทดลองนี้เป็นชุดข้อมูลมาตรฐานจาก UCI ที่ได้มาจากการเก็บข้อมูลจริง

การทดลองส่วนนี้เป็นการทดลองเพื่อศึกษาความสำคัญในการปรับค่าของตัวแปรแต่ละตัว ซึ่งแบ่งเป็นการทดลองย่อยๆ 3 การทดลอง ดังต่อไปนี้

- 1) ทดลองด้วย XCS มาตรฐาน (Conventional XCS)
- 2) ทดลองด้วย XCS ที่มีการปรับค่าอัตราการรับรู้ (XCS adaptive perception rate)
- 3) ทดลองด้วย XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold (XCS adaptive perception rate and GA threshold)

ในการทดลองแต่ละครั้งมีการกำหนดค่าสถานะในการทดลองเป็นดังต่อไปนี้ กำหนดจำนวนรอบข้อมูลที่ใช้ในการฝึกสอนเป็น 500 รอบชุดข้อมูล จำนวนประชากรของกฎมีค่าเป็น 6400 กฎ ค่าอัตราการเรียนรู้เริ่มต้นมีค่าเป็น 0.2 ค่า GA Threshold มีค่าเริ่มต้นเป็น 25 ค่า crossing-over threshold มีค่าเป็น 0.8 ค่า mutation threshold มีค่าเป็น 0.04 และกำหนดค่าตอบแทนสำหรับการจำแนกกลุ่มข้อมูล เป็น 2 ค่า ได้แก่ ค่าตอบแทนสูงสุด เป็น 1000 สำหรับกรณีที่ระบบสามารถจำแนกกลุ่มข้อมูลได้ถูกต้อง และค่าตอบแทนต่ำสุด เป็น 0 สำหรับกรณีที่ระบบจำแนกกลุ่มข้อมูลผิดพลาด

ตาราง 4.8 แสดงผลการจำแนกกลุ่มข้อมูลของ ชุดข้อมูลจริงที่มีกลุ่มข้อมูล 2 กลุ่ม โดยทดสอบกับ XCS 3 แบบ ได้แก่ 1) XCS มาตรฐาน 2) XCS ที่มีการปรับค่าอัตราการรับรู้ และ 3) XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold ผลลัพธ์ที่แสดงในตารางเป็นค่าเฉลี่ยจากผลการทดลองด้วย 10-fold cross-validation technique และ ค่าส่วนเบี่ยงเบนมาตรฐาน (standard deviation)

จากตาราง 4.8 แบ่งพิจารณาเป็น 2 กรณี ได้แก่ ชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลน้อย และชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลมาก ชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลน้อยจะคัดเลือกเฉพาะชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลตั้งแต่ 1 ถึง 5 ได้แก่ ชุดข้อมูล Breast Cancer Wisconsin, Diabetes, Tao Grid, Balance-Scale-dist2, Balance-scale-dist3, Glass-dist1, Glassdist2, Wine-dist1, Wine-dist2 และ Wine-dist3 และชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลน้อยจะคัดเลือกเฉพาะชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลตั้งแต่ 5 ขึ้นไป ได้แก่ ชุดข้อมูล Balance-scale-dist1, Glass-dist3 และ Glass-dist7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 4.8 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูลโดยวิธีการหาค่าเฉลี่ยเรขาคณิต และส่วนเบี่ยงเบนมาตรฐานของ XCS

Dataset Name	<i>ir</i>	Performance		
		Conventional XCS	XCS with adaptive perception rate	XCS with adaptive perception rate and GA
Breast Cancer Wisconsin	1.90 : 1	0.967 ± 0.012	0.968 ± 0.015	0.969 ± 0.014
Diabetes	1.87 : 1	0.681 ± 0.060	0.701 ± 0.067	0.673 ± 0.040
Heart-Statlog	1.25 : 1	0.804 ± 0.053	0.812 ± 0.088	0.822 ± 0.089
Tao Grid	1.00 : 1	0.901 ± 0.026	0.967 ± 0.011 ↑	0.965 ± 0.015 ↑
Balance-Scale-dist1	11.75 : 1	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
Balance-Scale-dist2	1.17 : 1	0.861 ± 0.049	0.843 ± 0.053	0.839 ± 0.055
Balance-Scale-dist3	1.17 : 1	0.840 ± 0.045	0.826 ± 0.042	0.828 ± 0.044
Glass-dist1	2.06 : 1	0.762 ± 0.106	0.847 ± 0.053 ↑	0.830 ± 0.103 ↑
Glass-dist2	1.82 : 1	0.759 ± 0.142	0.833 ± 0.103 ↑	0.802 ± 0.096 ↑
Glass-dist3	11.59 : 1	0.000 ± 0.000	0.277 ± 0.358	0.167 ± 0.360
Glass-dist7	6.38 : 1	0.867 ± 0.086	0.921 ± 0.091 ↑	0.937 ± 0.089 ↑
Wine-dist1	2.02 : 1	0.983 ± 0.041	0.991 ± 0.028	0.979 ± 0.046
Wine-dist2	1.50 : 1	0.970 ± 0.054	0.970 ± 0.054	0.958 ± 0.053
Wine-dist3	2.70 : 1	0.992 ± 0.017	0.996 ± 0.012	0.996 ± 0.012

พิจารณาที่ชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลน้อย จะพบว่า XCS ที่มีการปรับค่าอัตราการรับรู้ และ XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold สามารถจำแนกกลุ่มข้อมูลได้ดีกว่า XCS มาตรฐาน อย่างมีนัยสำคัญ ในการทดลองกับชุดข้อมูล Tao Grid, Glass-dist1 และ Glass-dist2 ส่วนชุดข้อมูลอื่นๆ ไม่มีความแตกต่างอย่างมีนัยสำคัญ

ในส่วนนี้จะเห็นว่าชุดข้อมูล Tao Grid เป็นชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลเป็น 1 แต่ XCS ที่มีการปรับค่าอัตราการรับรู้ และ XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold สามารถจำแนกกลุ่มข้อมูลได้ดีกว่า XCS มาตรฐานอย่างมีนัยสำคัญ เนื่องจากหัวข้อ 5.2.2 ได้อธิบายไว้ว่าชุดข้อมูล Tao Grid เป็นชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลพื้นที่ย่อย คิดเป็น 15.276 เท่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พิจารณาที่ชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลมาก จะพบว่า XCS ที่มีการปรับค่าอัตราการรับรู้ และ XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold สามารถจำแนกกลุ่มข้อมูลได้ดีกว่า XCS มาตรฐาน อย่างมีนัยสำคัญ ในการทดลองกับชุดข้อมูล Glass-dist7 ส่วนชุดข้อมูล Balance-Scale-dist1 ไม่มีความแตกต่างอย่างมีนัยสำคัญ

สำหรับการทดลองกับชุดข้อมูล Balance-scale-dist1 และ Glass-dist3 จะเห็นว่าประสิทธิภาพในการจำแนกและเรียนรู้ที่ได้จากตัวจำแนกประเภททั้ง 3 มีค่าน้อยมาก เนื่องจากชุดข้อมูล Balance-Scale-dist1 และ Glass-dist3 เป็นชุดข้อมูลที่มีอัตราความไม่สมดุลสูงมาก ทำให้ต้องการระยะเวลาในการเรียนรู้ที่นานมากขึ้น ตามรูป 5.7 ดังนั้นจึงทำการทดลองโดยเพิ่มจำนวนรอบการเรียนรู้เป็น 500 1000 และ 1500 สำหรับ ชุดข้อมูล Balance-scale-dist1 และ Glass-dist3 โดยมีผลลัพธ์ดังต่อไปนี้

ตาราง 4.9 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูลโดยวิธีการหาค่าเฉลี่ยเรขาคณิต และส่วนเบี่ยงเบนมาตรฐานของ XCS ในการทดลองกับชุดข้อมูล Balance-scale-dist1 และ Glass-dist3 เมื่อผ่านการเรียนรู้ที่ 500 1000 และ 1500 รอบ

Dataset Name	Performance		
	Conventional XCS	XCS with adaptive perception rate	XCS with adaptive perception rate and GA
Balance-Scale-dist1 500 iteration	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
Balance-Scale-dist1 1000 iteration	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
Balance-Scale-dist1 1500 iteration	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
Glass-dist3 500 iteration	0.000 ± 0.000	0.277 ± 0.358	0.167 ± 0.360
Glass-dist3 1000 iteration	0.000 ± 0.000	0.377 ± 0.408 ↑	0.477 ± 0.428 ↑
Glass-dist3 1500 iteration	0.000 ± 0.000	0.408 ± 0.442 ↑	0.477 ± 0.428 ↑

จากตาราง 4.9 จะพบว่าประสิทธิภาพในการจำแนกกลุ่มข้อมูลของ XCS ที่มีการปรับค่าอัตราการรับรู้ และ XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold เพิ่มขึ้นอย่างมีนัยสำคัญ เมื่อผ่านการเรียนรู้ที่ 1000 รอบขึ้นไป สำหรับชุดข้อมูล Balance-scale-dist1 XCS ไม่สามารถจำแนกกลุ่มเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ผ่านการอนุญาตจากผู้เกี่ยวข้อง หากต้องการข้อมูลเพิ่มเติม กรุณาติดต่อฝ่ายประชาสัมพันธ์ โทร. 0-2562-0000 หรือ 0-2562-0001

ข้อมูลได้เลย เนื่องจากชุดข้อมูล Balance-scale-dist1 มีการกระจายตัวของกลุ่มข้อมูล 1 น้อย และมีลักษณะเป็นเส้นตรง ทำให้ XCS ที่มีการเข้ารหัสในส่วนเงื่อนไขของกฎในรูปของ ขอบเขตบน-ของเขตล่าง ไม่สามารถสร้างกฎที่ครอบคลุม ข้อมูลของกลุ่มข้อมูล “1” ได้มากกว่า 1 จุดโดยไม่ครอบคลุมกลุ่มข้อมูล “0” ดังรูป 4.6



รูป 4.6 กราฟการกระจายตัวของชุดข้อมูล Balance-Scale-dist0

4.5.3 การทดลองกับชุดข้อมูลจริงที่มีกลุ่มข้อมูลตั้งแต่ 3 กลุ่มขึ้นไป

ในส่วนนี้เป็นการทดลองเพื่อทดสอบประสิทธิภาพในการจำแนกกลุ่มข้อมูลด้วย XCS สำหรับกลุ่มข้อมูลที่มีกลุ่มข้อมูลตั้งแต่ 3 กลุ่มขึ้นไป เนื่องจากชุดข้อมูลจริงโดยทั่วไปไม่ได้มีเพียง 2 กลุ่มข้อมูล ดังนั้นประสิทธิภาพในการจำแนกกลุ่มข้อมูลที่มีตั้งแต่ 3 กลุ่มขึ้นไปจึงมีความสำคัญในตัดสินประสิทธิภาพในการจำแนกโดยรวมของจักรกลเรียนรู้

ในการทดลองนี้เป็นการทดลองกับชุดข้อมูล 3 ชุด ได้แก่ ชุดข้อมูล Balance-Scale ชุดข้อมูล Glass และชุดข้อมูล Wine ชุดข้อมูล Balance-Scale และชุดข้อมูล Wine มีกลุ่มข้อมูลจำนวน 3 กลุ่ม ส่วนชุดข้อมูล Glass มีกลุ่มข้อมูลจำนวน 7 กลุ่ม แต่จะมีอยู่ 3 กลุ่มข้อมูลที่เป็นกลุ่มข้อมูลที่มีน้อยจนไม่สามารถวัดประสิทธิภาพด้วย 10 fold-cross validation technique ได้อย่างแม่นยำ ได้แก่ กลุ่มข้อมูลที่ 4 กลุ่มข้อมูลที่ 5 และกลุ่มข้อมูลที่ 6 ที่มีจำนวนข้อมูลเป็น 0 13 และ 9 ดังนั้นการวัด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประสิทธิภาพของชุดข้อมูล Glass จึงเลือกใช้เฉพาะกลุ่มข้อมูล 1 2 3 และ 7 ในการทดสอบ โดยได้ผลดังตารางต่อไปนี้

ตาราง 4.10 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูล และส่วนเบี่ยงเบนมาตรฐานของ XCS จากการทดลองกับชุดข้อมูลที่มีกลุ่มข้อมูลตั้งแต่ 3 กลุ่มขึ้นไป

Dataset Name	Performance		
	Conventional XCS	XCS adaptive with perception rate	XCS adaptive with perception rate and GA
Balance-Scale	0.717 ± 0.140	0.768 ± 0.076 ↑	0.797 ± 0.051 ↑
Glass	0.093 ± 0.294	0.293 ± 0.379	0.462 ± 0.405 ↑
Wine	0.963 ± 0.036	0.974 ± 0.039	0.974 ± 0.037

ตาราง 4.10 แสดงผลการจำแนกกลุ่มข้อมูลของ ชุดข้อมูลจริงที่มีกลุ่มข้อมูล 3 กลุ่ม โดยทดสอบกับ XCS 3 แบบ ได้แก่ 1) XCS มาตรฐาน 2) XCS ที่มีการปรับค่าอัตราการรับรู้ และ 3) XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold ผลลัพธ์ที่แสดงในตารางเป็นค่าเฉลี่ยจากผลการทดลองด้วย 10-fold cross-validation technique และ ค่าส่วนเบี่ยงเบนมาตรฐาน (standard deviation)

จากตาราง 4.10 จะพบว่า XCS ที่มีการปรับค่าอัตราการรับรู้ และ XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold สามารถจำแนกกลุ่มข้อมูลได้ดีกว่า XCS มาตรฐาน อย่างมีนัยสำคัญ ในการจำแนกทดสอบกับชุดข้อมูล Balance-Scale และ ไม่มีความแตกต่างอย่างมีนัยสำคัญในการทดสอบกับชุดข้อมูล Wine สำหรับการทดสอบกับชุดข้อมูล Glass พบว่า XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold สามารถจำแนกกลุ่มข้อมูลได้ดีกว่า XCS มาตรฐาน อย่างมีนัยสำคัญ

ในการทดลองกับชุดข้อมูลที่มี 2 กลุ่มข้อมูลพบว่า การทดสอบกับชุดข้อมูล Glass-dist3 ซึ่งเป็นชุดข้อมูลที่แปลงมาจากชุดข้อมูล Glass ดังนั้น จะได้ผลการจำแนกที่ดีหลังจากผ่านการเรียนรู้ที่ 1000 รอบชุดข้อมูลขึ้นไป จึงได้ทำการทดสอบกับชุดข้อมูล Glass เพิ่มเติม โดยกำหนดรอบการเรียนรู้เป็น 500 1000 และ 1500 โดยมีผลลัพธ์ดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 4.11 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูล และส่วนเบี่ยงเบนมาตรฐานของ XCS จากการทดลองกับชุดข้อมูล Glass ที่จำนวนรอบการเรียนรู้ 500 1000 และ 1500

Dataset Name	Performance		
	Conventional XCS	XCS adaptive perception rate	XCS adaptive perception rate and GA
Glass 500 iterations	0.093 ± 0.294	0.293 ± 0.379	0.462 ± 0.405 ↑
Glass 1000 iterations	0.093 ± 0.294	0.474 ± 0.415 ↑	0.529 ± 0.373 ↑
Glass 1500 iterations	0.156 ± 0.331	0.474 ± 0.415 ↑	0.544 ± 0.383 ↑

จากตาราง 4.11 พบว่า XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold มีประสิทธิภาพในการจำแนกกลุ่มข้อมูลของชุดข้อมูล Glass ได้ดีกว่า XCS ที่ต้นแบบ อย่างมีนัยสำคัญ หลังจากผ่านการเรียนรู้ที่ 500 รอบชุดข้อมูลขึ้นไป และ XCS ที่มีการปรับค่าอัตราการรับรู้ มีประสิทธิภาพในการจำแนกกลุ่มข้อมูลของชุดข้อมูล Glass ได้ดีกว่า XCS มาตรฐาน อย่างมีนัยสำคัญ หลังจากผ่านการเรียนรู้ที่ 1000 รอบชุดข้อมูลขึ้นไป

4.6 การเปรียบเทียบผลการทดลอง

ในหัวข้อนี้นำเสนอการทดลองเพื่อเปรียบเทียบ XCS ที่มีตัวแปรปรับค่าได้ กับจักรกลเรียนรู้ชนิดต่างๆ ได้แก่ โดยเปรียบเทียบ XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold กับ ต้นไม้การตัดสินใจชนิด C4.5 (Decision tree: C4.5) ที่มีการนำเสนอองค์ความรู้ในรูปของขอบเขตการตัดสินใจ ซึ่งมีการนำเสนอองค์ความรู้คล้ายคลึงกับ XCS ที่เป็น lower-upper bound และทดสอบกับระบบโครงข่ายประสาทเทียม (Neural Network) และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ในลักษณะของ linear separate โดยจักรกลเรียนรู้ทั้งหมดที่ใช้ในการเปรียบเทียบนี้เป็นจักรกลเรียนรู้มาตรฐานที่รวมเข้ากับชุดทดลองการทำเหมืองข้อมูล (Data mining) ที่ชื่อว่า WEKA การทดลองเพื่อเปรียบเทียบนี้แบ่งออกเป็น 2 การทดลอง ได้แก่ การทดลองกับชุดข้อมูลที่มีกลุ่มข้อมูล 2 กลุ่ม และการทดลองกับชุดข้อมูลที่มีกลุ่มข้อมูลมากกว่า 2 กลุ่ม

จากตาราง 4.12 และ 4.13 จะเห็นได้ว่า XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold กับ ต้นไม้การตัดสินใจชนิด C4.5 มีประสิทธิภาพในการจำแนกกลุ่มข้อมูลใกล้เคียงกันมากกว่าระบบโครงข่ายประสาทเทียม และซัพพอร์ตเวกเตอร์แมชชีน เนื่องจากรูปแบบในการนำเสนอองค์ความรู้ เนื่องจาก XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold และ ต้นไม้การตัดสินใจ มี

การนำเสนอองค์ความรู้ในรูปแบบของกฎ ส่วนระบบโครงข่ายประสาทเทียม และซัพพอร์ตเวกเตอร์แมชชีน มีการนำเสนอองค์ความรู้ในรูปแบบของการแบ่งกลุ่มแบบเชิงเส้น (linear separate)

ตาราง 4.12 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูลและส่วนเบี่ยงเบนมาตรฐานของ XCS จากการทดลองกับชุดข้อมูลที่มีกลุ่มข้อมูล 2 กลุ่ม

Dataset Name	Performance			
	XCS with adaptive perception rate and GA	C4.5	Neural Network	SVM
Breast Cancer Wisconsin	0.969 ± 0.014	0.960 ± 0.022	0.954 ± 0.024	0.965 ± 0.016
Diabetes	0.673 ± 0.040	0.709 ± 0.052	0.728 ± 0.067	0.691 ± 0.068
Heart-Statlog	0.822 ± 0.089	0.789 ± 0.078	0.820 ± 0.084	0.829 ± 0.071
Tao Grid	0.965 ± 0.015	0.954 ± 0.021	$0.883 \pm 0.039 \downarrow$	$0.840 \pm 0.027 \downarrow$
Balance-Scale-dist0	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
Balance-Scale-dist1	0.839 ± 0.055	0.818 ± 0.043	$0.943 \pm 0.027 \uparrow$	$0.924 \pm 0.031 \uparrow$
Balance-Scale-dist2	0.828 ± 0.044	0.824 ± 0.067	$0.959 \pm 0.020 \uparrow$	$0.926 \pm 0.024 \uparrow$
Glass-dist1	0.830 ± 0.103	0.840 ± 0.128	$0.782 \pm 0.072 \downarrow$	$0.319 \pm 0.228 \downarrow$
Glass-dist2	0.802 ± 0.096	$0.749 \pm 0.161 \downarrow$	$0.663 \pm 0.181 \downarrow$	$0.000 \pm 0.000 \downarrow$
Glass-dist3	0.167 ± 0.360	$0.508 \pm 0.453 \uparrow$	0.000 ± 0.000	0.000 ± 0.000
Glass-dist7	0.937 ± 0.089	$0.875 \pm 0.143 \downarrow$	0.918 ± 0.088	0.900 ± 0.089
Wine-dist0	0.979 ± 0.046	0.957 ± 0.045	0.996 ± 0.013	0.983 ± 0.037
Wine-dist1	0.958 ± 0.053	0.923 ± 0.102	0.968 ± 0.035	0.977 ± 0.052
Wine-dist2	0.996 ± 0.012	0.950 ± 0.060	0.986 ± 0.034	0.992 ± 0.017

โดยจะเห็นได้ว่า XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold กับ ค้นไม่การตัดสินใจชนิด C4.5 สามารถจำแนกกลุ่มข้อมูลของชุดข้อมูล Tao ได้ดีกว่า ระบบโครงข่ายประสาทเทียมและซัพพอร์ตเวกเตอร์แมชชีน เนื่องจากชุดข้อมูล Tao มีการกระจายตัวของกลุ่มข้อมูลย่อยแฝงอยู่ในอีกกลุ่มข้อมูล ทำให้เหมาะกับจักรกลเรียนรู้ที่มีการนำเสนอองค์ความรู้ในรูปแบบของกฎมากกว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 4.13 ค่าความถูกต้องในการจำแนกกลุ่มข้อมูลของ และส่วนเบี่ยงเบนมาตรฐานของ XCS จากการทดสอบกับชุดข้อมูลที่มีกลุ่มข้อมูลมากกว่า 2 กลุ่ม

Dataset Name	Performance			
	XCS with adaptive perception rate and GA	C4.5	NN	SVM
Balance-Scale	0.797 ± 0.051	$0.902 \pm 0.062 \uparrow$	$0.906 \pm 0.061 \uparrow$	0.746 ± 0.131
Glass	0.462 ± 0.405	$0.000 \pm 0.000 \downarrow$	$0.000 \pm 0.000 \downarrow$	$0.000 \pm 0.000 \downarrow$
Wine	0.974 ± 0.037	$0.939 \pm 0.054 \downarrow$	0.978 ± 0.029	0.985 ± 0.024

แต่สำหรับชุดข้อมูล Balance-Scale ที่เหมาะสำหรับการแบ่งกลุ่มข้อมูลแบบเชิงเส้น ระบบโครงข่ายประสาทเทียมและซัพพอร์ตเวกเตอร์แมชชีน จะสามารถจำแนกกลุ่มข้อมูลของชุดข้อมูลได้ดีกว่า XCS ที่มีค่าปรับค่าอัตราการเรียนรู้และ GA threshold กับ ดัชนีการตัดสินใจชนิด C4.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ชุดข้อมูลที่ไม่สมดุล คือ ชุดข้อมูลที่มีจำนวนของข้อมูลในกลุ่มใดกลุ่มหนึ่งมากหรือน้อยกว่าจำนวนข้อมูลในกลุ่มอื่นหลายๆ ความไม่สมดุลของกลุ่มข้อมูลที่เกิดขึ้นนี้ส่งผลให้การเรียนรู้ของจักรกลเรียนรู้โน้มเอียงเข้าหากกลุ่มข้อมูลที่มีข้อมูลจำนวนมาก จนไม่สามารถเรียนรู้กลุ่มข้อมูลที่มีจำนวนข้อมูลน้อยๆ ได้ เนื่องจากจักรกลเรียนรู้โดยส่วนใหญ่แล้วพัฒนาขึ้นมาโดยมีสมมติฐานว่า “จำนวนข้อมูลแต่ละกลุ่มในชุดข้อมูลฝึกสอน (data training) มีจำนวนเท่าๆ กันในแต่ละกลุ่ม”

XCS เป็นระบบจำแนกประเภทแบบเรียนรู้ได้ที่ได้รับการยกย่องว่ามีประสิทธิภาพมากที่สุดระบบหนึ่ง XCS เป็นระบบที่ผสมผสานการเรียนรู้แบบเสริมกำลังและเจเนติกอัลกอริทึมเข้าไว้ด้วยกัน เพื่อสร้างเซตของกฎที่สามารถบรรยายความรู้ที่ได้จากชุดข้อมูลที่น่ามาฝึกสอนให้มนุษย์เข้าใจได้ ระบบนี้มีข้อดี คือ การนำเสนอในรูปแบบ เงื่อนไข-การกระทำ โดย ซึ่งง่ายต่อการทำความเข้าใจของมนุษย์ แต่อย่างไรก็ตามการศึกษา XCS กับชุดข้อมูลที่ไม่สมดุลในระดับที่สูงมากๆ ดังนั้นจึงจำเป็นต้องมีการพัฒนาให้ XCS สามารถใช้งานได้กับชุดข้อมูลที่ไม่สมดุลในระดับที่สูงมากๆ ได้

ในงานวิจัยฉบับนี้นำเสนอวิธีการพัฒนา XCS ให้สามารถเรียนรู้ชุดข้อมูลที่ไม่สมดุลได้ โดยการปรับค่าอัตราการรับและค่า GA threshold ให้สอดคล้องกับอัตราส่วนความไม่สมดุลของกลุ่มข้อมูลที่เกิดขึ้นระหว่างการเรียนรู้

จากการทดลองกับชุดข้อมูลจริงพบว่า การปรับค่าอัตราการรับและค่า GA threshold ด้วยวิธีการที่นำเสนอสามารถเพิ่มประสิทธิภาพของ XCS ในการจำแนกกลุ่มข้อมูลของชุดข้อมูลจริงที่มีความไม่สมดุลได้มากๆ อย่างมีประสิทธิภาพ และนอกจากวิธีการที่นำเสนอก็สามารถเพิ่มประสิทธิภาพในการจำแนกกลุ่มของชุดข้อมูลที่มีอัตราส่วนความไม่สมดุลไม่มาก แต่มีการกระจายตัวของข้อมูลออกเป็นกลุ่มย่อยหลายๆ กลุ่มย่อยได้ดีขึ้นอย่างมีนัยสำคัญ

จากการเปรียบเทียบกับจักรกลเรียนรู้ประเภทอื่นพบว่า XCS ที่มีการปรับค่าอัตราการรับและ GA threshold สามารถจำแนกกลุ่มข้อมูลที่มีการกระจายตัวเป็นกลุ่มก้อนได้ดีกว่า กลุ่มข้อมูลที่มีการกระจายตัวเป็นเส้นทแยงมุม

จากการทดลองและเปรียบเทียบกับจักรกลเรียนรู้อื่นๆ พบว่า XCS เหมาะกับชุดข้อมูล binary มากที่สุด โดยจะมีประสิทธิภาพในการจำแนกกลุ่มข้อมูลที่ไม่สมดุลมากๆ ได้ สำหรับชุดข้อมูลจริง XCS จะเหมาะกับชุดข้อมูลที่มีการกระจายตัวเป็นกลุ่มข้อมูลย่อยๆ สลับกัน ในการทดสอบกับชุดข้อมูลลักษณะนี้ XCS จะสามารถจำแนกกลุ่มข้อมูลได้ดีกว่าจักรกลเรียนรู้ชนิดอื่นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 ข้อเสนอแนะ

XCS ที่มีการปรับค่าอัตราการรับรู้และ GA threshold ยังจำแนกกลุ่มข้อมูลที่มีลักษณะเป็นเส้นทแยงมุมได้ไม่ดี เนื่องจากส่วนเงื่อนไขของกฎในระบบจำแนกประเภทมีการเข้ารหัสเป็นขอบเขตล่าง-ขอบเขตบน ซึ่งมีลักษณะเป็นกรอบสี่เหลี่ยม ดังนั้นการปรับแก้การเข้ารหัสในส่วนเงื่อนไขให้มีลักษณะเป็นวงรี จะทำให้ XCS รองรับชุดข้อมูลที่มีความหลากหลายมากยิ่งขึ้น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

- [1] J. H. Holland, “**Adaptation**”. In Rosen & Snell (eds), *Progress in Theoretical Biology*, 4. Plenum, 1976.
- [2] S. W. Wilson, “**ZCS: Zeroth Level Classifier System**”, *Evolutionary Computation*, vol. 2, no. 1, 1994, page 1-18.
- [3] S. W. Wilson, “**Classifier Fitness Based on Accuracy**”, *Evolutionary Computation*, vol. 3, no. 2, 1995, page 149-176.
- [4] L. Bull, “**A Simple Accuracy-based Learning Classifier System**”, *University of the West of England, Bristol*, 2003.
- [5] M. V. Butz and S. W. Wilson, “**An Algorithmic Description of XCS**”, *Journal of Soft Computing* 6, 2002, pages 144-153.
- [6] J. Taeho and J. Nathalie, “**Class imbalances versus small disjuncts**”, *ACM SIGKDD Explorations Newsletter*, vol. 6, 2004, pages 40-49.
- [7] Q. Zhou, and ed. al., “**Using an improved C4.5 for imbalanced dataset of intrusion**”, In *Proceedings of the 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST '06)*. ACM, New York, USA., 2006.
- [8] N. V. Chawla, “**C4.5 and imbalanced data sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure**”, *Workshop on Learning from Imbalanced Datasets II, ICML, Washington DC*, 2003.
- [9] R. Akbani, S. Kwek, J. Nathalie, “**Applying Support Vector Machines to imbalanced datasets**”, In *Proceedings of the 15th European Conference on Machine Learning (ECML)*, 2004, pages 39-50.
- [10] J. Nathalie and S. Shaju. “**The class imbalance problem: A systematic study**”. *Intelligent Data Analysis*. vol. 6, 2002, pages 429-449.

- [11] A. P. Orriols and E. Bernardó, “**Bounding XCS’s Parameters for unbalanced datasets**”, *GECCO*, Washington, USA, July 8-12, 2006.
- [12] A.W. Moore, “**Cross-validation for detecting and preventing over fitting**”, www.cs.cmu.edu/~awm/tutorials, 2001.
- [13] X. Llorá and J.M. Garrell. “**Inducing partially-defined instances with Evolutionary Algorithms**”, In *Proceedings of the 18th International Conference on Machine Learning* (To Appear), 2001.
- [14] Albert Orriols Puig and Ester Bernardó-Mansilla, “*The Class Imbalance Problem in Learning Classifier Systems: A Preliminary Study*,” *Proceedings of 2005 workshops on Genetic and Evolutionary Computing*, Washington D.C., USA, 2005, pages 74-78.
- [15] Q. Gu, L. Zhu and Z. Cai, “**Evaluation Measures of the Classification Performance of imbalanced data sets**”, *COMPUTATIONAL INTELLIGENCE AND INTELLIGENT SYSTEMS*, 2009, vol. 51, part 9, pages 461-471.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

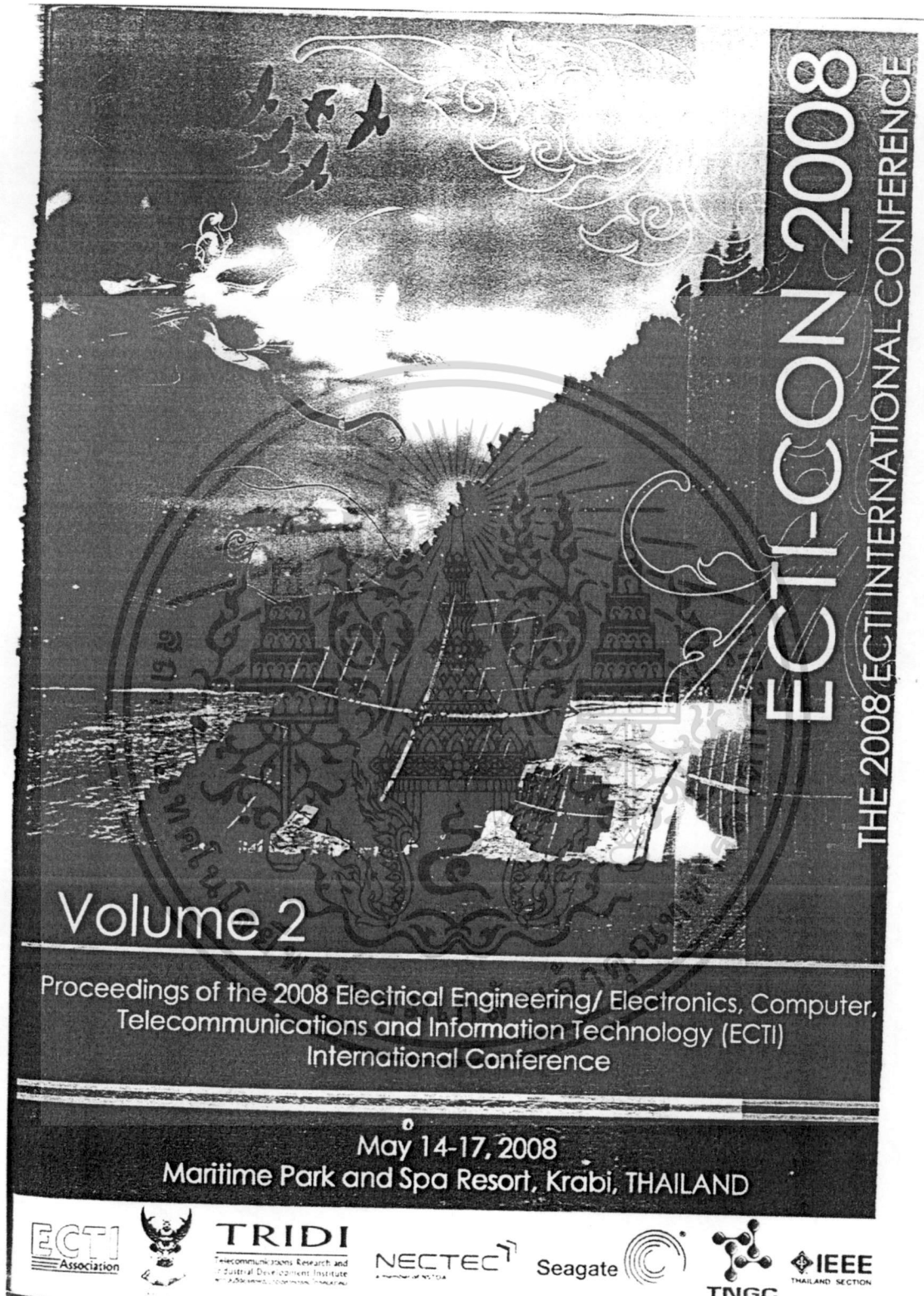
ภาคผนวก



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Using Accuracy-Based Learning Classifier Systems for Imbalance Datasets

Somchai Udomthanapong*, Kreangsak Tamee†, and Ouen Pinngern‡

Department of Computer Engineering, Faculty of Engineering
 Research Center for Communication and Information Technology (ReCCIT)
 King Mongkut's Institute of Technology Ladkrabang Bangkok, Thailand 10520
 Email: u.somchai@gmail.com*, kreangsakt@hotmail.com†, kpouen@kmitl.ac.th‡

Abstract—XCS is one of the most powerful learning classifier systems. It combines reinforcement learning and genetic algorithm to create a set of rules representing the extracted knowledge from dataset. The main advantage of this system is to provide rule-based models that represent human-readable patterns. However, not too much public have yet been studied in imbalance dataset.

In this paper, we propose a novel technique to develop XCS deal with imbalance dataset. The proposed technique uses adaptive perception rate for each rule to provide balance learning between major and minor class. The experiment show that the propose technique can classify all level of imbalance classes on the well-know Boolean logic benchmark task - multiplexer problem.

I. INTRODUCTION

Learning Classifier Systems (LCSs) [1] are machine learning that are hybrid between Genetic Algorithm and Reinforcement Learning. LCS implements a rule based approach to reinforcement learning. The systems represent the set of agents knowledge as a population of condition-action-prediction rules that are called classifiers. Recently, LCSs and their derived system, i.e. ZCS [2], XCS [3],[4], YCS [5], UCS [6],[7], etc., have been enhanced to solve challenging machine learning problems and implement in previous applications [8],[9]. However, learning with imbalance datasets are not widely focusing on [6],[7],[10].

Imbalance datasets are the datasets that contain one of the classes is represented by a very small number of examples compared to the other classes. This case occurs in several domains such as fraud detection, oil spills in satellite images, failures in manufacturing process, etc. Many machine learners assume a uniform distribution of classes, so that they may suffer from biases toward the majority class when they are exposed to high levels of classes imbalance. Most researches in imbalance classes they use SVM [11], [12] and ID3 [13].

Researches in the field of XCS have few studies influence of imbalance. Orriols and Bernadó analyzed behavior of XCS with imbalance datasets [10],[14],[6],[7]. They showed XCS with standard setting parameter, it implemented as described in [4], cannot learn and classify high imbalance datasets and presented guidelines to set XCS's parameters based-on imbalance ratio of datasets.

In this paper we proposed a method to develop XCS for imbalance datasets based-on a guideline of Orriols and Bernadó. The developed XCS is able to classify 100% at

imbalance level from $i = 1$ to $i = 10$ of datasets, under bounding that Orriols and Bernadó had proposed, for the description see section IV.

The paper is organized as follows. Section II describes Accuracy-based Learning Classifier System: XCS. Section III explains the standard datasets for XCS that is called multiplexer problem. Section IV shows a developing Orriols and Bernadó algorithm for imbalance datasets. Section V present a result of each algorithm. Section VI provide conclusion and future work.

II. DESCRIPTION OF ACCURACY-BASED LEARNING CLASSIFIER SYSTEM: XCS

An Accuracy-based Learning Classifier System (XCS) was introduced by Wilson[3]. It was developed from the Holland's traditional system. Moreover, today it has been recognized as the best and most popular LCS. The most importance components of XCS are knowledge representation, parameter updates and discovery component.

XCS represents the agent knowledge as a population set $[P]$ of rule or classifier. Each rule consist of condition, action and prediction parameter. In condition part $C \in \{0, 1, \#\}$ and action $A \in \{a1, a2, \dots, an\}$ specifies the action proposes. In prediction parameters use to predict a received reward from environment and identify strange of classifier such as prediction payoff (p), prediction error (ϵ), Fitness (F), niche size estimate (as).

The XCS work as the follow: When the system receives input message (in binary form) it will create a match set $[M]$ to contain the classifiers from population set $[P]$ that can match this input state. If the match set $[M]$ is empty the covering operator is used to generate the classifiers that can match current input. After that, the system will generate prediction array from the match set $[M]$ by averaging of prediction for every proposed action by using the following equation:

$$x = \frac{\sum_{cl.a=a \wedge cl \in [M]} cl.p \times cl.F}{\sum_{cl.a=a \wedge cl \in [M]} cl.F} \quad (1)$$

In selection an action to take to environment, XCS selects action by using roulette wheel of average the prediction payoff in exploration and selects maximum prediction payoff in

exploitation. After that an action set $[A]$ is generated from classifier in match set that contain selected action is performed.

After received reward (R) from the environment, all classifiers in the action set $[A]$ are updated prediction parameters using Q-learning technique of reinforcement learning. The updated equation is computed as:

$$p = p + \beta(R - p) \quad (2)$$

$$\varepsilon = \varepsilon + \beta(|R - p| - \varepsilon) \quad (3)$$

$$as = as + \beta(|A| - as) \quad (4)$$

where β is a learning rate. $\beta \in (0, 1)$. Fitness updating is based on accuracy that we can calculate from the following equation:

$$k = \begin{cases} \alpha(\varepsilon/\varepsilon_0)^{-v} & \text{if } \varepsilon > \varepsilon_0 \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

$$k' = \frac{k}{\sum_{j \in [A]} k} \quad (6)$$

where parameter ε_0 ($\varepsilon_0 > 0$) is a threshold for accepted error. If some classifiers have ε less than ε_0 then this classifier predicts true and set k equal 1. For α ($\alpha \in (0, 1)$) and v ($v > 1$) is constant for calculate accuracy. After that we can calculate fitness using the following equation:

$$F = F + \beta(k' - F) \quad (7)$$

Discovery component, in XCS has two operators for discovery rule. First covering operator: If the match set $[M]$ is empty it will generate new classifier for current message input. Every condition bit the system will randomly select between current input bit and '#' and random one of the proposed action. Then the system initials prediction parameter's values by default. Second genetic algorithm: after the system updating and averaging of time stamp (ts) more than threshold ($\theta_{G,A}$), XCS will select two classifiers from action set $[A]$ by using roulette wheel of fitness for run genetic operators. For initial values of prediction parameters, we use averaging value from parents. The new rule that comes from discovery operations should accept the current input.

III. MULTIPLEXER PROBLEM

In XCS, a multiplexer problem is standard datasets, used to training and testing performance of systems and were introduced by Wilson [3]. Multiplexer problem are represented in binary string of concatenate address and data. We can calculate length of those string by $n + 2^n$ where n is length of address, and separated into two class, '0' or '1'. For training and testing system, multiplexer datasets are generated by random between '0' and '1' for each bit.

Orriols and Bernadó present imbalance level as $i \in \{1, 10\}$ and define each level based-on concurrent ratio of major class '0' with minor class '1' as 2^i . We can generate imbalance data sets for training and testing XCS by setting ratio of class '0' (major class) with class '1' (minor class).

IV. XCS FOR IMBALANCE DATASETS

Orriols and Bernadó proposed that there are some parameters that has an effect with XCS's leaning. They analyzed that if we run XCS parameter $\varepsilon_0 = 0$ and maximum reward $R_{max} = 1000$ it should classify minority class correctly when $1 \leq ir \leq 1998$, the parameter imbalance ratio (ir) is a ratio between number of instance from major class N_{maj} and number of instance from minor class N_{min} . When they tested conventional XCS on imbalance datasets that create from multiplexer problem at level $i = 0$ to $i = 9$ it can classify minority class at level $i = 0$ to $i = 4$ result show as figure 1. So that unless ε_0 and β there are some problems to have effect with XCS's learning.

According from this study, Orriols and Bernadó tried to developed XCS for imbalance datasets by analyzed XCS's parameters that have influence with accuracy of classification for minor class. They proposed that the learning rate β should be set as ratio of frequencies occurring of the least class f_{min} per frequencies occurring of the most class f_{max} and parameter $\theta_{G,A}$ should be set greater than the delay between two examples of the minority class, i.e.

$$\beta = k_1 \cdot \frac{f_{min}}{f_{max}} \quad (8)$$

$$\theta_{G,A} = k_2 \cdot \frac{1}{f_{min}} \quad (9)$$

where k_1 is arbitrary constant for tuning the learning rate, $k_1 \in (0, 1)$ and k_2 is constant for tuning $\theta_{G,A}$, $k_2 > 1$.

However, XCS with the tuning follow Orriols and Bernadó proposed can lean at low and medium level of imbalance classes but it cannot learn the height rate of imbalance ratio, see in figure 2. Furthermore, it very slow to lean at medium levels of imbalance classes and unable to classify the minority class in high imbalance levels. It might be effect of the setting learning rate close to zero, for example at imbalance levels $i = 9$ the learning rate is set as $0.2/29 = 0.000390$ and imbalance level at $i = 10$ the learning rate is set $0.2/210 = 0.000195$. From this reasons show that a little learning rate is not good way to develop.

We develop XCS for imbalance datasets based on assumption that a rule will have high number of wrong prediction if it is an overgeneralize rule. We should set its learning rate lower than rule have never predict incorrect. We implement an adaptive parameter, it is called perception rate (δ) into update equation for p and ε_0 . For adaptive perception rate, it will be adapted if it predicts incorrectly, i.e. parameter p not equal to reward.

$$p = p + \beta \cdot \delta \cdot (R - p) \quad (10)$$

$$\delta = (\zeta)^{\omega} \quad (11)$$

where ω is wrong prediction number of classifier and ζ is decrease rate. $\zeta \in (0, 1)$.

V. EXPERIMENTS AND RESULTS

In this section, we compare the three version of XCSs: conventional XCS, Orriols and Bernadó's XCS, XCS with adaptive perception rate (our proposed method) on 11-bit multiplexer problem which create as imbalance class at different levels $i = 1$ to 10. For each level the system was run 10 times with different seeds. In addition, every experiment we trained XCS on boolean multiplexer in 5,000,000 iterations. For each iteration, if the system classify correctly it receive a reward 1000; otherwise it receives a reward 0. Parameters in each experiment are set as the following.

For conventional XCS used the same parameter setting and by Wilson [4] as follow: population size (N) = 2000, $\beta = 0.2$, $\alpha = 0.1$, $\varepsilon_0 = 1$, $v = 5$, $\lambda = 0.8$, $\mu = 0.04$ and $\theta_{G,A} = 25$.

The parameter setting for Orriols and Bernadó's XCS similar to Orriols and Bernadó [10] as follow: $k_1 = 0.2$ and $k_2 = 5$.

For our proposed XCS, we use the same parameters setting as previous and $\zeta = 0.8$.

Figure 1 show the true negative (TN) rate, i.e., the ratio of correct classifications for class 0, and the true positive (TP) rate, i.e., the ratio of correct classifications for class 1, obtained by XCS with imbalance levels from $i = 1$ to $i = 10$. Figure 1(a) shows that XCS can learn negative class quickly for all imbalance levels, the graphs reach to 100% when the learning iteration less than 10000. However the figure 1(b) shows XCS can learn positive class only first 5 levels, can reach to 100% after 10,000, 20,000, 40,000, 80,000 and 300,000 respectively. While the curve at level $i = 6$ stabilizing at 80% after 250,000 iterations. For higher imbalance datasets levels, from $i = 7$ to $i = 10$, the system cannot classify the positive class.

Figure 2 shows the true negative rate and the true positive rate, Orriols and Bernadó's XCS. We can see that they can improve TP rate for level $i = 8$. However, the TP rate at $i = 9$ still perform poor.

Figure 3, XCS with adaptive perception rate. We can see that perform significantly better then conventional XCS and Orriols and Bernadó's XCS both on imbalance levels $i = 9$ and $i = 10$.

In term of learning speed, if we focus on imbalance datasets level $i = 5$. The conventional XCS are able to classify a positive 100% after 330,000 iterations learning and Orriols and Bernadó's XCS can classify classify a positive 100% after learn 240,000 iterations and by XCS with adaptive perception rate can classify classify a positive 100% after only 120,000 iterations. It means that the XCS with adaptive perception rate can improve leaning speed of XCS.

VI. CONCLUSION

This paper proposed a version of XCS, adaptive perception rate. We used adaptive perception rate for balance leaning between major and minor classes.

The proposed method, XCS with adaptive perception rate, was test on multiplexer problem with difference imbalance levels from $i = 1$ to $i = 10$. It is shown that the true positive rate of the proposed method is better than conventional XCS

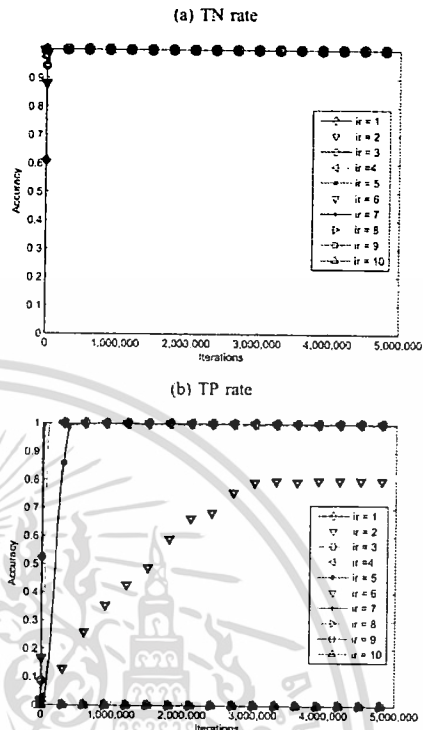


Fig. 1. TN rate (a) and TP rate (b) in 11-bit multiplexer problem from imbalance levels 1-10. Parameter set same conventional XCS.

and Orriols and Bernadó's XCS on tested imbalance levels of 11-bit multiplexer problem. In term of learning speed, we showed that our proposed method can improve learning speed of XCS.

For further research, we would like to study effects of noise to our system and also we would like to analyze on other artificial problems and real-world datasets.

REFERENCES

- [1] J. H. Holland, *Adaptation*, In Rosen & Snell (eds), Progress in Theoretical Biology, 4, Plenum, 1976.
- [2] S. W. Wilson, *ZCS: Zeroth Level Classifier System*, Evolutionary Computation, vol. 2, no. 1, page 1-18, 1994.
- [3] S. W. Wilson, *Classifier Fitness Based on Accuracy*, Evolutionary Computation, vol. 3, no. 2, page 149-176, 1995.
- [4] M. V. Butz and S. W. Wilson, *An Algorithmic Description of XCS*, Journal of Soft Computing 6, pages 144-153, 2002.
- [5] L. Bull, *A Simple Accuracy-based Learning Classifier System*, University of the West of England, Bristol, 2003.

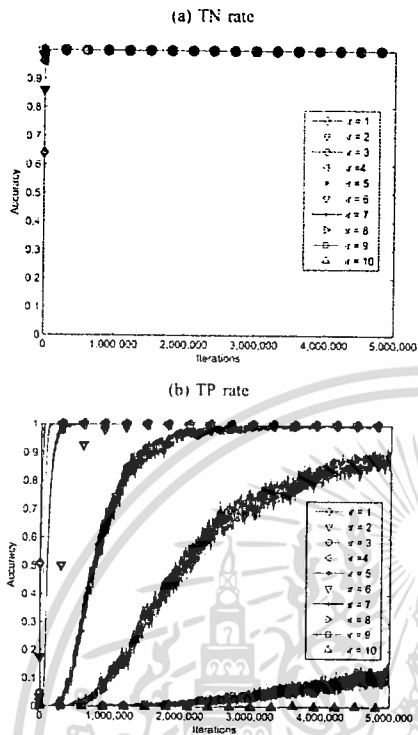


Fig. 2. TN rate (a) and TP rate (b) in 11-bit multiplexer problem from imbalance levels 1-10. Parameters set according to Orríols and Bernadó.

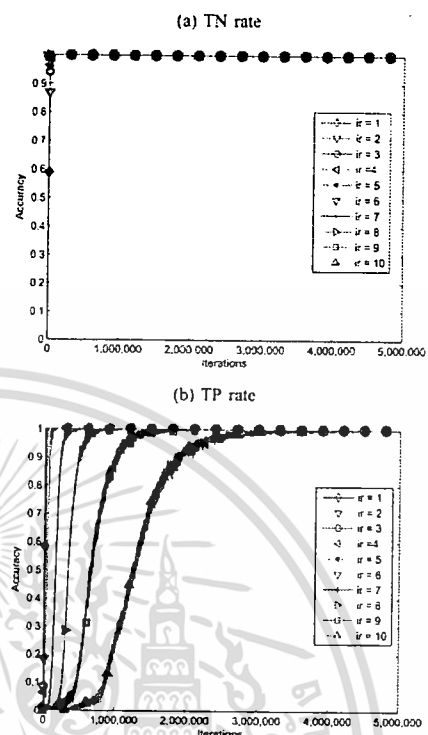



Fig. 3. TN rate (a) and TP rate (b) in 11-bit multiplexer problem from imbalance levels 1-10. XCS's Parameter adapt as follow presented algorithm.

- [6] A. Orríols and E. Bernadó, *The Class Imbalance Problem in Learning Classifier Systems: A Preliminary Study*, Proceedings of 2005 workshops on Genetic and Evolutionary Computing, Washington D.C., USA, pages 74-78, 2005.
- [7] A. Orríols and E. Bernadó, *Class Imbalance Problem in UCS Classifier System: Fitness Adaptation*, IEEE Transactions on Evolutionary Computation, vol. 1, pages 604-611, 2005.
- [8] A. J. Bagnall and G. C. Cawley, *Learning Classifier Systems for Data Mining: A Comparison of ACS with Other Classifiers for the Forest Cover Data Set*, Proceedings of the International Joint Conference on Neural Network, vol. 3, pages 1802-1807, 2003.
- [9] L. Bull, (ed.), *Applications of Learning Classifier Systems*, Springer, 2004.
- [10] A. Orríols and E. Bernadó, *Bounding ACS's Parameters for Unbalanced Datasets*, GECCO, Washington, USA, July 8-12, 2006.
- [11] Y. Y. Ou, et al., *A Study of Supervised Learning with Multivariate Analysis on Unbalanced Datasets*, International Joint Conference on Neural Networks, pages 2201-2205.
- [12] Q. Tao, et al., *Posterior Probabilities Support Vector Machines for Unbalanced Data*, IEEE Transactions on Neural Network, vol. 16, no. 6, pages 1561-1573, 2006.
- [13] J. Auer and R. Hall, *Investigating ID3-Induced Rules from Low-Dimensional Data Cleaned by Complete Case Analysis*, AI 2004: Advances in Artificial Intelligence, Springer Berlin, pages 414-424, 2004.
- [14] H. H. Dam, et al., *Can Evolutionary Computation Handle Large Datasets? A Study into Network Intrusion Detection*, AI 2005: Advances in Artificial Intelligence, Springer Berlin, pages 1092-1095, 2005.


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



NCSEC 2008


NCSEC 2008
The 12th National Computer Science and Engineering Conference

Past, Present, and Future of Research in Thai Computing
 November 20th - 21st, 2008


Sponsored by:


NECTEC
A member of NECTA


UniNet


WD Western Digital

Organized by:


ELECTICAL
Engineering Institute

Department of Electrical Engineering
 King Mongkut's University of Technology
 North Bangkok
 Power | Control | Communication | Computer

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การพัฒนาประสิทธิภาพของระบบ XCS สำหรับปัญหาความไม่สมดุลของกลุ่มข้อมูล
Improving XCS performance for Imbalance Problem

ศรัชัย คุณชนาทพงศ์ วิทยาภาส ไรญุรักษ์ ศรีนิล เกรียงศักดิ์ เคมีย์ เอื้อน ปิ่นนิม

ห้องวิจัยศาสตร์ข้อมูล สำนักวิจัยการสื่อสารและเทคโนโลยีสารสนเทศ

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ถ.คลองจungk เขตลาดกระบัง กรุงเทพฯ 10520 โทรศัพท์ : 0-2737-3000

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง

ถ. รามคำแหง แขวงหัวหมาก เขตบางกะปิ กรุงเทพฯ 10240 โทรศัพท์ : 0-2310-8022

E-mail: s9060705@kmitl.ac.th, s9060022@kmitl.ac.th, s8060020@kmitl.ac.th, kreangsakt@hotmail.com, ouen@ru.ac.th

บทคัดย่อ

ปัญหาความไม่สมดุลกันของกลุ่มข้อมูลเป็นหัวข้อที่ได้รับ
ความสนใจเนื่องจากเป็นปัญหาสำหรับการเรียนรู้ของระบบเรียนรู้ ระบบ
เรียนรู้โดยส่วนใหญ่จะให้ค่าประสิทธิภาพต่ำเมื่อเรียนรู้กับปัญหาประเภท
นี้. ในงานวิจัยนี้เราได้ศึกษาผลกระทบของปัญหาความไม่สมดุลกันของ
กลุ่มข้อมูลที่มีต่อระบบ XCS. ซึ่งเป็นระบบเรียนรู้จำแนกประเภทชนิด
หนึ่งที่มีประสิทธิภาพ รวมทั้งนำเสนอวิธีการพัฒนาประสิทธิภาพของ
ระบบ โดยนำเสนอทราบดีกว่าในการรับรู้ที่สามารถปรับเปลี่ยนได้
สำหรับกฎแต่ละกฎเพื่อทำให้เกิดสมดุลในการเรียนรู้ระหว่างกลุ่มที่มี
ข้อมูลจำนวนมากและกลุ่มที่มีข้อมูลจำนวนสมาชิกน้อย ผลการ
ทดลองแสดงให้เห็นว่าวิธีการที่นำเสนอสามารถเอาชนะความไม่สมดุลได้
ทุกระดับในปัญหามัลติเทสทซ์เซตอร์ ซึ่งเป็นปัญหาที่นิยมใช้ทดสอบ
ระบบ โดยทำการทดลองทั้งการสร้างข้อมูลแบบเคลื่อนที่บนระบบ 10-
fold cross validation.

Abstract

In recent years, learning with imbalanced data sets receives
more and more attentions. Almost learning systems perform poor when
tackle with this problem. In this paper, we study the effect of class
imbalance problem in XCS which is one of the most powerful learning
classifier systems and proposed a method to improve the performance.
The proposed method uses adaptive perception rate for each rule to
provide balance learning between majority and minority class. The
experiments indicate that the propose method can classify 10 levels of
imbalance classes on the well-known 11-bits Boolean logic benchmark
task - multiplexer problem - generated from both on-the-fly and 10-fold

cross validation.

Keywords: learning classifier systems, imbalance,
classification, font style, blank line

1. Introduction

A Learning Classifier System (LCS) [1] is a rule-based
machine learning system that combines between Reinforcement
Learning and Genetic Algorithm. The reinforcement learning
component is responsible for adjusting the strength of rules in the
system according to some reward obtained from the environment. The
genetic algorithm acts as an innovation discovery component which
plays a key role in discovering novel and potentially useful rules. The
advantages of this approach are rule-based representation, which make
it easy to interpret by human; online learning ability, where the system
is updated after the presentation of each data instance; robustness due to
the use of genetic algorithm.

In 1995, XCS [8] emerged as a type of LCS where a
classifier's fitness depends on the prediction of its expected payoff
instead of the actual prediction itself as in traditional LCS which
worked better on both reinforcement and supervised learning problems.
Recently, XCS has been widely applied to classification and data mining
problems [2]. An investigation [9] on 15 UCI datasets showed that
XCS's performance is at least as good as traditional machine learning
techniques. In [10] study compared XCS with C4.5, the Naive Bayes
classifier, PART, the instance based learning algorithm with one and
three nearest neighbor settings, and the support vector machine on 42



datasets. Statistical tests of significance showed that XCS outperforms other algorithms in some datasets.

Among them, mining imbalance data has been emphasized as one of most interested topic in XCS. A class imbalance problem is the datasets that contain one of the classes is represented by a very small number of examples compared to the other classes. This case occurs in several domains such as fraud detection, oil spills in satellite images, failures in manufacturing process, etc. Many machine learners assume a uniform distribution of classes, so that they may suffer from biases toward the majority class when they are exposed to high levels of class imbalance. Researches in the field of XCS have few studies influence of imbalance. Oriols and Bernadó analyzed behavior of XCS with imbalance datasets [5]. They showed XCS with standard setting parameter, if implemented as described in [6], cannot learn and classify high imbalance datasets and presented guidelines to set XCS's parameters based on imbalance ratio of datasets.

In this paper we proposed a method to develop XCS for imbalance datasets based on a guideline of Oriols and Bernadó. The developed XCS is able to classify 100% at imbalance level from $i = 1$ to $i = 10$ of datasets, under bounding that Oriols and Bernadó had proposed.

The rest of this paper is organized as follow. Section 2 provide introduction of XCS for data mining. Section 3 describes the well-know Boolean logic benchmark task - multiplexer problem. Section 4 illustrates how to improve XCS for imbalance problem. Section 5 shows the experimental result. Conclusions are drawn and future work is discussed in Section 6.

2. Description of Accuracy-based Learning Classifier

System: XCS

An Accuracy-based Learning Classifier System (XCS) was introduced by Wilson[7]. It was developed from the Holland's traditional system. Moreover, today it has been recognized as the best and most popular LCS. The most importance components of XCS are knowledge representation, parameter updates and discovery component.

XCS represents the agent knowledge as a population set $[P]$ of rule or classifier. Each rule consists of condition, action and

prediction parameter. In condition part $C \in \{0, 1, \#\}^L$ and action $A \in \{a_1, a_2, \dots, a_n\}$ specifies the action proposes. In prediction parameters, there are three principal parameters; prediction payoff (p) which estimates the payoff for that classifier if classifier is selected to action, prediction error (ε) which estimates the average error made in the payoff predictions, fitness (F) which can calculate based on error.

The XCS work as the follow: First, XCS finds a match set $[M]$ with all classifiers in the population $[P]$ whose condition matches the incoming instance. If the match set $[M]$ is empty then the covering operator is used to generate the classifiers that match current input. After that, the system generate prediction array from the match set $[M]$ by averaging of prediction for every possible action using the following equation:

$$p_j = p_j + \beta(R - p_j) \quad (1)$$

In action selection, XCS selects action by using roulette wheel of average the prediction payoff for exploration phase and selects maximum prediction payoff for exploitation phase. After that an action set $[A]$ is generated from classifier in match set $[M]$ that have the same action as the chosen action:

After received reward (R) from the environment, all parameters of classifiers in the action set $[A]$ are updated using Q-learning technique of reinforcement learning. The updated equation is computed as:

$$p_j = p_j + \beta(R - p_j) \quad (2)$$

$$\varepsilon_j = \varepsilon_j + \beta(|R - p_j| - \varepsilon_j) \quad (3)$$

$$as_j = as_j + \beta(|A| - as_j) \quad (4)$$

where β is a learning rate, $\beta \in (0, 1)$. Fitness updating is based on accuracy that we can calculate form the following equation:

$$K_j = \begin{cases} \alpha(\varepsilon_j / \varepsilon_0)^{-r} & \text{if } \varepsilon_j > \varepsilon_0 \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

$$K_j = \frac{K_j}{\sum_{j \in [A]} K_j} \quad (6)$$



where parameter ε_0 ($\varepsilon_0 > 0$) is a threshold for accepted error. If some classifiers have ε less than ε then this classifier predicts true and set k equal 1. For α ($\alpha \in (0, 1)$) and ν ($\nu > 1$) is constant for calculate accuracy. After that we can calculate fitness using the following equation:

$$F_j = F_j + \beta(\kappa_j - F_j) \quad (7)$$

Discovery component in XCS has two parts. First covering operator, when match set $[M]$ is empty it will generate new classifier for current message input. Every condition bit in rule will randomly select between current input bit and ψ -and random one of possible action. Then the system initials prediction parameter's values by default value. Second, GA is invoke when time stamp (ts) more than threshold (θ_{GA}). If GA is activated, XCS will select two classifiers from action set $[A]$ by using roulette wheel on their fitness. If the population size reaches a predefined limit, some classifiers are removed by voting within the population.

3. Multiplexer Problem

The multiplexer problem is traditionally studied in the LCS literature due to its interesting function properties by Wilson in 1985 [7]. Multiplexer data are defined as binary strings of length $l = k + 2^k$ that consist of 2 part; the first k bits as an address and last 2^k bits as data. The class value is determined by the value of data that refer by address. For example, in the 6-multiplexer ($l = 2 + 2^2 = 6$), the class value for the input string 100010 is 1, because the "address", 10, indexes bit 2 of the remaining four bits. 11-multiplexer and 20-multiplexer are the next more-complicated multiplexers. The corresponding expression for 11-multiplexer has eight terms each consisting of four factors; for 20-multiplexer there are 16 terms of five factors each.

We used 11-multiplexer in our experiment. We divided our experiment into two parts. Firstly, on-the-fly experiment, the train and test data is directly generated by function. The imbalance multiplexer [6] permits to control the imbalance complexity of the multiplexer by undersampling the class labeled as '1'. Secondly, validation experiment, we create offline 11-multiplexer by generate all of possible value and

the imbalance complexity is controlled by oversampling the class labeled as '0' contain the proportion of imbalance ratio which is a ratio between number of instance from major class N_{cm} , and number of instance from minor class N_{cm} . Then training set is divided into 10 subsets in a stratified sample. A 10-fold cross validation is used with a 9:1 train/validation proportion. The system is trained on each training set in batch mode and then tested on validation set. Each experiment was run 10 times with different random seeds.

4. XCS for Imbalance Datasets

Orriols and Bernadó proposed that there are some parameters that have an effect with XCS's learning. They analyzed that if we run XCS with parameter $\varepsilon_0=0$ and maximum reward $R_{max} = 1000$ then it should classify minority class correctly when $1 < ir < 1998$. When they tested conventional XCS on imbalance datasets that created from multiplexer problem at level $i=0$ to $i=9$. They can classify minority class at level $i = 0$ to $i = 4$ result show as figure 1. So that unless ε_0 and β there are some problems to have effect with XCS's learning.

We improve XCS for imbalance datasets based on assumption that the imbalance ratio in each classifier is not equal. From the assumption, we propose a parameter to expect an imbalance ratio that classifier cover to update equation for p and ε_0 as following:

$$p_j = p_j + \psi \cdot \beta(R - p_j) \quad (8)$$

$$\varepsilon_j = \varepsilon_j + \psi \cdot \beta(|R - p_j| - \varepsilon_j) \quad (9)$$

where ψ is a parameter for estimate imbalance ratio, it called adaptive perception rate. To estimate an imbalance ratio in each classifier, we calculate by ratio between number of correct classify and number of incorrect classify as follow equation:

$$\psi = \begin{cases} 1 & \text{if } c = 0, c' = 0 \\ c/c' & \text{if } c > c' \\ c'/c & \text{if } c < c' \end{cases} \quad (10)$$

where c is number of correct classify and c' is number of incorrect classify.

5. Experiments and Results

5.1 Experiment on on-the-fly 11-multiplexer

In this section, we compare the three versions of XCSs: conventional XCS, Orriols and Bernadó's XCS, proposed XCS on 11-bit multiplexer problem which create as imbalance class at different levels $i = 1$ to $i = 10$. In addition, every experiment we trained XCS on Boolean multiplexer in 5,000,000 iterations. In each iteration, if the system correctly classify then it will receive a reward 1000; otherwise it will receive a reward 0. The parameter setting for three XCSs same as Wilson [7] as follow: population size (N) = 2000, $\beta = 0.2$, $\alpha = 0.1$, $\epsilon_0 = 1$, $v = 5$, $\chi = 0.8$, and $\mu = 0.04$. The performance from exploit trials on is recorded, using a 100-point average, average over ten run.

Figure 1 shows the true negative (TN) rate, i.e., the ratio of correct classifications for class '0', and the true positive (TP) rate, i.e., the ratio of correct classifications for class 1, obtained by XCS with imbalance levels from $i = 1$ to $i = 10$. Figure 1(a) shows that XCS can learn negative class quickly for all imbalance levels, the graphs reach to 100% when the learning iteration less than 10000. However the figure 1(b) shows XCS can learn positive class only first 5 levels, can reach to 100% after 10,000, 20,000, 40,000, 80,000 and 300,000 respectively. While the curve at level $i = 6$ stabilizing at 80% after 250,000 iterations. For higher imbalance datasets levels, from $i = 7$ to $i = 10$, the system cannot classify the positive class.

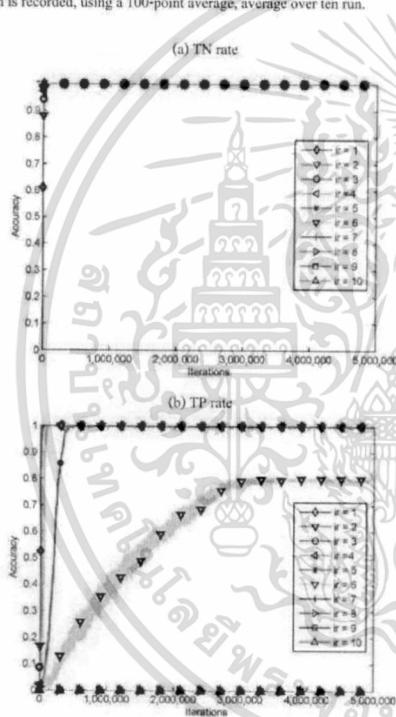


Fig. 1. TN rate (a) and TP rate (b) in 11-bit multiplexer problem from imbalance levels 1-10. Parameters set according to conventional XCS

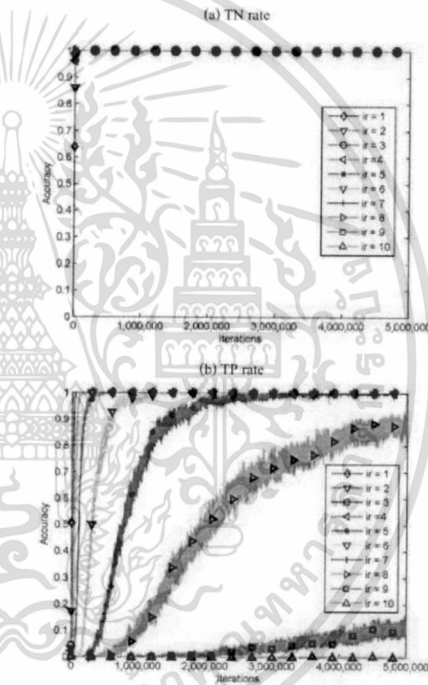


Fig. 2. TN rate (a) and TP rate (b) in 11-bit multiplexer problem from imbalance levels 1-10. Parameters set according to Orriols and Bernadó.



NCSEC
2008

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Figure 2 shows the true negative rate and the true positive rate, Orriols and Bernadó's XCS. We can see that they can improve TP rate for level $i = 8$. However, the TP rate at $i = 9$ still perform poor.

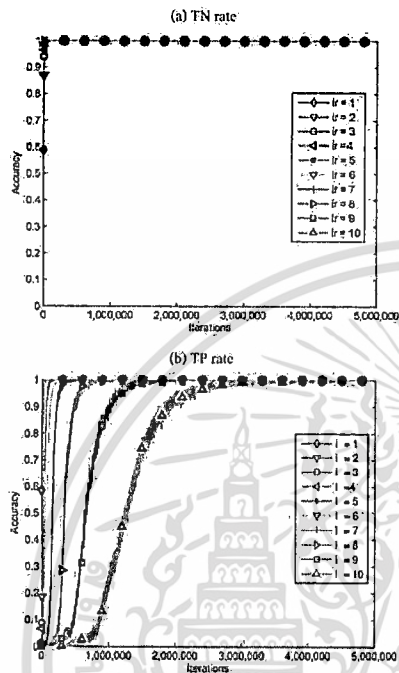


Fig. 3. TN rate (a) and TP rate (b) of XCS with propose method for 11-bit multiplexer problem from imbalance levels 1-10.

Figure 3, XCS with adaptive perception rate. We can see that perform significantly better than conventional XCS and Orriols and Bernadó's XCS both on imbalance levels $i = 9$ and $i = 10$.

In term of learning speed, if we focus on imbalance datasets level $i = 5$. The conventional XCS is able to classify a positive 100% after 330,000 iterations learning and Orriols and Bernadó's XCS can classify a positive 100% after learn 240,000 iterations and by XCS with adaptive perception rate can classify a positive 100% after only 120,000 iterations. It means that XCS with adaptive perception rate can improve learning speed of XCS.

5.2 Experiment on 10-fold cross validation. 11-bit multiplexer

In this section, we compare the three versions of XCSs on offline 11-multiplexer. Differences that are statistically significant at a significance level of 0.05 are denoted by a \uparrow if other propose is better than conventional XCS and by a \downarrow if other propose is worse than conventional XCS, \blacklozenge if other propose is better than Orriols and Bernadó's and by a \blacklozenge if other propose is worse than Orriols and Bernadó's XCS.

Table 1. Comparison TP rate between three XCSs on 11-bits multiplexer dataset from imbalance levels 6-10.

Method	Level	TP rate
Conventional XCS	$i = 6$	0 ± 0.00
	$i = 7$	0 ± 0.00
	$i = 8$	0 ± 0.00
	$i = 9$	0 ± 0.00
	$i = 10$	0 ± 0.00
Orriols and Bernadó's XCS	$i = 6$	$1.00 \pm 0.00 \uparrow$
	$i = 7$	$1.00 \pm 0.00 \uparrow$
	$i = 8$	$0.98 \pm 0.13 \uparrow$
	$i = 9$	$0.54 \pm 0.17 \uparrow$
	$i = 10$	$0.02 \pm 0.09 \uparrow$
XCS with propose method	$i = 6$	$1.00 \pm 0.00 \uparrow$
	$i = 7$	$1.00 \pm 0.00 \uparrow$
	$i = 8$	$1.00 \pm 0.00 \uparrow \blacklozenge$
	$i = 9$	$1.00 \pm 0.00 \uparrow \blacklozenge$
	$i = 10$	$0.90 \pm 0.12 \uparrow \blacklozenge$

Table 1. presents the mean and the standard deviation (over 10 runs) of the predictive accuracy in different system on the offline 11-bit multiplexer data sets with imbalance dataset from level $i = 6$ to $i = 10$. The result shows that the accuracy of Orriols and Bernadó's XCS is significantly better than conventional XCS in imbalance level $i = 6$ to $i = 9$. However, our proposed XCS is significantly better than conventional XCS in imbalance level $i = 6$ to $i = 10$ and also

significantly better than Orriols and Bernadó's XCS in imbalance level $i = 8$ to $i = 10$.

6. Conclusions

This paper proposed an improved version of XCS. We used adaptive perception rate for balance learning between major and minor classes.

The proposed method, XCS with adaptive perception rate, was tested on multiplexer problem with different imbalance levels from $i = 1$ to $i = 10$ both in on-the-fly and 10-fold cross validation. It is shown that the true positive rate (TP) of the proposed method is better than conventional XCS and Orriols and Bernadó's XCS on tested imbalance levels of 11-bit multiplexer problem. In term of learning speed, we showed that our proposed method can improve learning speed of XCS. For further research, we would like to study effects of noise to our system and also we would like to analyze on other artificial problems and real-world datasets.

7. Acknowledgements

Work reported in this paper was funded by the Research center Communication and Information Technology.

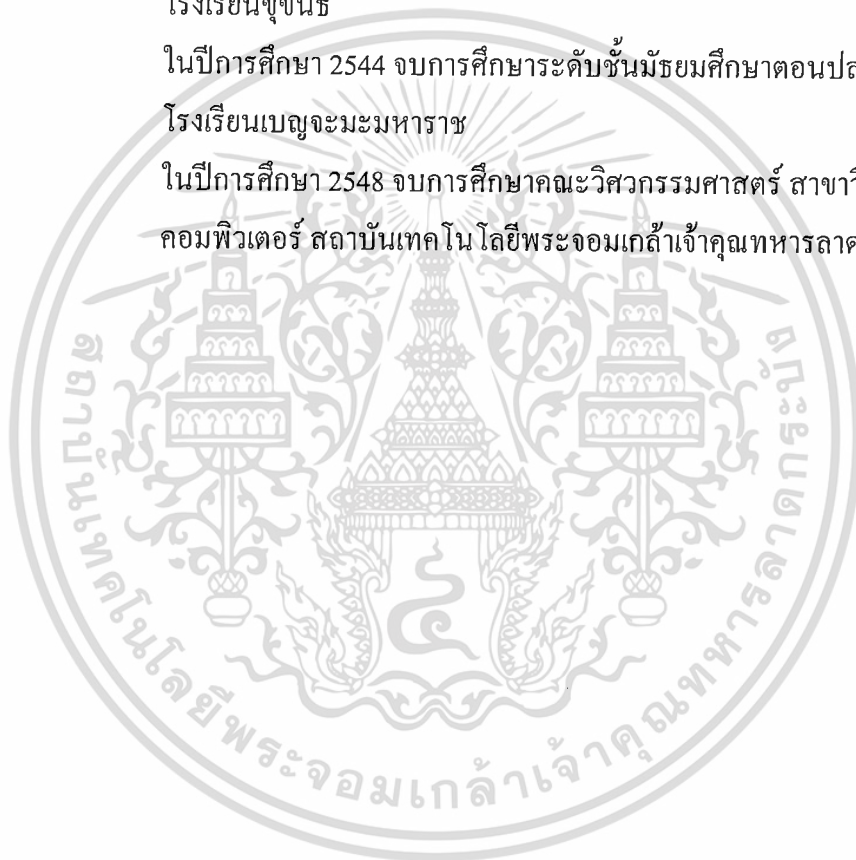
References

- [1] J.H. Holland, "Adaptation", In Rosen & Snell (eds), *Progress in Theoretical Biology*, 4, Plenum, 1976.
- [2] L. Bull, E. Bernadó-Mansilla and J. Holmes, "Learning Classifier System in Data Mining", *Springer*, 2008.
- [3] Q. Tao, et al, "Posterior Probability Support Vector Machines for Unbalanced Data", *IEEE Transactions on Neural Networks*, Vol. 16; No. 6, pp. 1561-1573, 2006.
- [4] J. Auer and R. Hall, "Investigating ID3-Induced Rules from Low-Dimensional Data Cleaned by Complete Case Analysis", *AI 2004: Advances in Artificial Intelligence*, Springer Berlin, pp. 414-424, 2004.
- [5] A. Orriols and E. Bernadó, "The Class Imbalance Problem in Learning Classifier Systems: A Preliminary Study", *Proceedings of 2005 workshops on Genetic and Evolutionary Computing*, Washington D.C., USA, pp. 74-78, 2005.
- [6] A. Orriols and E. Bernadó, "Bounding XCS's Parameters for Unbalanced Datasets", *GECCO*, Washington, USA, July 8-12, 2006.
- [7] M. V. Butz and S. W. Wilson, "An Algorithmic Description of XCS", *Journal of Soft Computing* 6, pp. 144-153, 2002.
- [8] S. W. Wilson, "Classifier Fitness Based on Accuracy", *Evolutionary Computation*, Vol. 3, No. 2, page 149-176, 1995.
- [9] Ester Bernadó-Mansilla, Xavier Llor'a, and Josep M. Garrell-Guiu. XCS and GALE: a comparative study of two learning classifier systems with six other learning algorithms on classification tasks. *IWLS-2001*, pages 337-341, 2001.
- [10] Martin V. Butz, *Rule-based Evolutionary Online Learning Systems: Learning Bounds, Classification, and Prediction*. PhD thesis, University of Illinois at Urbana-Champaign, 2004.



ประวัติผู้เขียน

ชื่อ-นามสกุล	นายศรชัย อุดมธนาพงศ์
วันเดือนปีเกิด	วันที่ 20 สิงหาคม 2526
ที่อยู่	727 หมู่ 6 ถนนสุขุมวิท อำเภอสุขุมวิท จังหวัดศรีสะเกษ
ประวัติการศึกษา	ในปีการศึกษา 2541 จบการศึกษาระดับชั้นมัธยมศึกษาตอนต้นจาก โรงเรียนสุขุมวิท ในปีการศึกษา 2544 จบการศึกษาระดับชั้นมัธยมศึกษาตอนปลายจาก โรงเรียนเบญจมาฆะมหาธาต ในปีการศึกษา 2548 จบการศึกษาคณะวิศวกรรมศาสตร์ สาขาวิชา คอมพิวเตอร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้