การทำนายการยกเลิกบริการของลูกค้าโดยใช้เทคนิคการเรียนรู้เสริมกำลัง

น.ส.เมธาวี ปัญจสุชาติ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Customer Churn Prediction using Reinforcement Learning Technique

Miss Maythawee Panjasuchat

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2019

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การทำนายการยกเลิกบริการของลูกค้าโดยใช้เทคนิคการเรียนรู้เสริมกำลัง
โดย	น.ส.เมธาวี ปัญจสุชาติ
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.ญาใจ ลิ่มปิยะกรณ์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.ญาใจ ลิ่มปิยะกรณ์)

..... กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.ภาสกร อภิรักษ์วรพินิต)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เมธาวี ปัญจสุชาติ : การทำนายการยกเลิกบริการของลูกค้าโดยใช้เทคนิคการเรียนรู้เสริมกำลัง. (Customer Churn Prediction using Reinforcement Learning Technique) อ.ที่ปรึกษาหลัก : รศ. ดร.ญาใจ ลิ่มปิยะกรณ์

การทำนายการยกเลิกบริการของลูกค้าเป็นหนึ่งในความท้าทายที่ยิ่งใหญ่สำหรับการดำเนินงานธุรกิจในปัจจุบัน เนื่องจากการสูญเสียลูกค้าย่อมส่งผลโดยตรงต่อชื่อเสียง แผนการเงิน และการเติบโตขององค์กร พฤติกรรมลูกค้าอาจเปลี่ยนแปลงไปจากเหตุปัจจัยที่ไม่สามารถควบคุมได้หรือสถานการณ์ที่ไม่อาจคาดคิด ส่งผลกระทบต่อแพตเทิร์นของข้อมูลที่เปลี่ยนไป ซึ่งอาจส่งผลลบต่อความสามารถการทำนายของตัวจำแนกประเภทที่สร้างจากเทคนิคการเรียนรู้แบบมีการชี้นำ ซึ่งเป็นการเรียนรู้ที่ไม่โต้ตอบ งานวิจัยนี้จึงนำเสนอการประยุกต์ใช้เทคนิคการเรียนรู้เสริมกำลังสำหรับการทำนายการยกเลิกบริการของลูกค้าในธุรกิจโทรคมนาคม แบบจำลองดีคิวเอ็นและโพลิซีเกรเดียนต์ได้ถูกพัฒนาและปรับใช้เพื่อการเรียนรู้บนชุดข้อมูลการยกเลิกบริการของลูกค้าที่ใช้สำหรับงานการจำแนกประเภท ชุดข้อมูลอีกชุดหนึ่งได้ถูกสร้างขึ้นจากชุดข้อมูลเดิมโดยการเพิ่มตัวอย่างที่แสดงถึงพฤติกรรมลูกค้าที่เปลี่ยนแปลงไป สมรรถนะของตัวจำแนกประเภทที่เลือกเปรียบเทียบกับดีคิวเอ็นและโพลิซีเกรเดียนต์ถูกประเมินด้วยค่าตัววัดทั้งสี่ประกอบด้วย ความแม่นยำ ความเที่ยงตรง ความครบถ้วน และเอฟวัน ผลการทดลองพบว่า เมื่อเปรียบเทียบกับตัวจำแนกประเภทที่เลือกคือ เอกซ์จีบูสต์ แรนดอมฟอเรสต์ และ เคเอ็นเอ็น ตัวเรียนรู้เชิงรุก ดีคิวเอ็นและโพลิซีเกรเดียนต์ มีสมรรถนะที่เหนือกว่าทั้งสองสถานการณ์ กล่าวคือ เมื่อทดสอบด้วยชุดข้อมูลตั้งต้น และเมื่อชุดข้อมูลได้ขยายตัวเพิ่มขึ้นและวิวัฒนาการจากการเกิดขึ้นของแพตเทิร์นใหม่

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่อนิสิต .....

ปีการศึกษา 2562 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

6170951021 : MAJOR COMPUTER SCIENCE

KEYWORD: Reinforcement Learning, churn prediction, active learning, supervised learning dataset

Maythawee Panjasuchat : Customer Churn Prediction using Reinforcement Learning Technique. Advisor: Assoc. Prof. Yachai Limpiyakorn, Ph.D.

Customer churn prediction is one of the biggest challenges for business nowadays, since the loss of customers directly affects the organization's reputation, financial and growth plans. Customer behaviors may change due to any uncontrollable factors or unexpected circumstances, resulting in changing patterns of data. This may aggravate predictability of the classifiers generated from supervised learning technique considered as Passive learning. This research has thus proposed applying the technique of reinforcement learning for customer churn prediction in telecommunication business. The models of Deep Q Network (DQN) and Policy Gradient have been implemented and adapted for learning on the selected customer churn dataset used for classification tasks. Another dataset was created from the original set with additional samples reflecting customer behavior changes. The performance of the selected classifiers, compared to DQN and Policy Gradient, has been evaluated with four measures: accuracy, precision, recall, and F1. The experimental results showed that, compared to the selected classifiers: XGBoost, Random forest, and kNN, the Active learners, DQN and Policy Gradient outperformed in both scenarios, that is, when testing with the original dataset and when the dataset was enlarged and evolving with emerging new patterns.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Field of Study: Computer Science Student's Signature .....

Academic Year: 2019 Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดีด้วยความอนุเคราะห์จากรองศาสตราจารย์ ดร.ญาใจ ลิ่มปิยะกรณ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้สละเวลาอันมีค่ามาให้ความรู้ คำปรึกษา รวมถึงตรวจสอบและแก้ไขปัญหาข้อผิดพลาดต่าง ๆ ที่เกิดขึ้น ตลอดจนการกำกับดูแลและคอยติดตามความก้าวหน้า ทำให้งานวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ และอาจารย์ ดร.ภาสกร อภิรักษ์วรพินิต กรรมการสอบวิทยานิพนธ์ ที่กรุณาเสียสละเวลาอันมีค่า ให้คำแนะนำที่เป็นประโยชน์ในการทำวิจัย ตลอดจนการตรวจสอบและแก้ไขวิทยานิพนธ์ฉบับนี้

ขอขอบพระคุณมารดา และญาติพี่น้องที่ให้การสนับสนุน ความรัก และคอยให้กำลังใจเป็นอย่างดีเสมอมา

ขอขอบคุณเพื่อน ๆ พี่ ๆ น้อง ๆ ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ทุกคน ที่คอยช่วยเหลือและให้คำปรึกษา แบ่งปันความรู้ และความคิดเห็นในด้านต่าง ๆ รวมถึงคอยให้กำลังใจตลอดระยะเวลาที่ผ่านมา

สุดท้ายนี้ขอขอบคุณผู้ที่เกี่ยวข้องทุกคนที่ไม่ได้กล่าวลงนามไว้ ณ ที่นี้ ที่มีส่วนทำให้วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดี ผู้วิจัยหวังเป็นอย่างยิ่งว่าวิทยานิพนธ์ฉบับนี้จะเป็นประโยชน์ต่อผู้ที่สนใจ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เมธาวี ปัญจสุชาติ

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ค

บทคัดย่อภาษาอังกฤษ ..... ง

กิตติกรรมประกาศ ..... จ

สารบัญ ..... ฉ

สารบัญตาราง ..... ซ

สารบัญรูปภาพ ..... ฌ

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 2

1.3 ขอบเขตการดำเนินงาน 2

1.4 ขั้นตอนการวิจัย 3

1.5 ประโยชน์ที่คาดว่าจะได้รับ 3

1.6 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์ 3

1.7 ผลงานที่ได้รับการตีพิมพ์จากวิทยานิพนธ์ 4

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 5

2.1 ทฤษฎีที่เกี่ยวข้อง 5

2.1.1 กระบวนการตัดสินใจแบบมาร์คอฟ (Markov Decision Process) 5

2.1.2 การเรียนรู้เสริมกำลัง 5

2.1.3 แบบจำลอง Deep Q Network (DQN) 9

2.1.4 แบบจำลอง Policy gradient 10

2.2 งานวิจัยที่เกี่ยวข้อง 11

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2.1 งานวิจัยเกี่ยวกับการทำนายลักษณะลูกค้าที่จะยกเลิกบริการ 11
2.2.2 งานวิจัยเกี่ยวกับการใช้เทคนิคการเรียนรู้เสริมกำลังกับโจทย์การเรียนรู้มีการชี้นำ 13
บทที่ 3 แนวคิดและวิธีวิจัย 16
3.1 ข้อมูลที่ใช้ในงานวิจัยและการประมวลผลข้อมูลเบื้องต้น 16
3.1.1 การทำความสะอาดข้อมูล (data cleaning) 16
3.1.2 การแปลงข้อมูล (Data transformation) 17
3.1.3 การคัดเลือกตัวแปร (Feature selection) 17
3.1.4 เตรียมชุดข้อมูลสำหรับการทดลองข้อมูลมีแพตเทิร์นเปลี่ยนไปจากเดิม 19
3.2 โครงสร้างของแบบจำลองการเรียนรู้เสริมกำลัง 21
3.2.1 ขั้นตอนการเรียนรู้ของแบบจำลองการเรียนรู้เสริมกำลัง DQN 21
3.2.2 ขั้นตอนการเรียนรู้ของแบบจำลองการเรียนรู้เสริมกำลัง Policy gradient 24
3.3 แนวทางการวัดสมรรถนะของแบบจำลอง 25
บทที่ 4 การทดลองและผลการทดลอง 28
4.1 การปรับช่วงข้อมูลของตัวแปร 28
4.2 เปรียบเทียบสมรรถนะกับแบบจำลองการเรียนรู้ที่มีการชี้นำอื่น ๆ 28
4.2.1 แบบจำลองแรนดอมฟอเรสต์ 29
4.2.2 แบบจำลองเอกซ์จีบูสต์ 30
4.2.3 แบบจำลองวิธีหาเพื่อนบ้านใกล้ที่สุด k ตัว 30
4.3 ผลการทดลอง 31
บทที่ 5 สรุปผลการวิจัย 35
5.1 สรุปผลการวิจัย 35
5.2 แนวทางงานวิจัยในอนาคต 36
บรรณานุกรม 37
ประวัติผู้เขียน 39

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

หน้า

ตารางที่ 1 ชื่อตัวแปร ความหมาย และคะแนนความสำคัญของ 21 ตัวแปร 18

ตารางที่ 2 อัตราส่วนของตัวแปรป้ายในชุดข้อมูลสอน และข้อมูลทดสอบกรณีแพตเทิร์นของข้อมูลเปลี่ยนแปลงไปจากเดิม 21

ตารางที่ 3 คอนฟิวชันเมทริกซ์ที่ใช้จำแนกแบบสองคลาส 25

ตารางที่ 4 การปรับจูนค่าไฮเปอร์พารามิเตอร์ในแบบจำลองแรนดอมฟอเรสต์ 29

ตารางที่ 5 การปรับจูนค่าไฮเปอร์พารามิเตอร์ในแบบจำลองเอกซ์จีบูสต์ 30

ตารางที่ 6 การปรับจูนค่าไฮเปอร์พารามิเตอร์ในแบบจำลองวิธีหาเพื่อนบ้านใกล้ที่สุด k ตัว 31

ตารางที่ 7 สรุปเปรียบเทียบสมรรถนะระหว่างแบบจำลองด้วยค่าความแม่นยำ ความเที่ยงตรง ความครบถ้วน และเอฟวันของข้อมูลชุดที่หนึ่ง และข้อมูลชุดที่มีแพตเทิร์นเปลี่ยนไป 32

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญรูปภาพ

หน้า

ภาพที่ 1 การสื่อสารระหว่างระบบ และสิ่งแวดล้อมในกระบวนการตัดสินใจแบบมาร์คอฟ [3] 6

ภาพที่ 2 ขั้นตอนวิธีการเรียนรู้ของแบบจำลอง DQN [2] 10

ภาพที่ 3 ขั้นตอนวิธีการเรียนรู้ของแบบจำลอง Policy gradient [2] 11

ภาพที่ 4 อัตราส่วนของตัวแปรป้ายระหว่างข้อมูลสอนและข้อมูลทดสอบของ NSL-KDD [2] 14

ภาพที่ 5 อัตราส่วนของตัวแปรป้ายระหว่างข้อมูลสอนและข้อมูลทดสอบของ AWID [2] 14

ภาพที่ 6 ผลการทดลองสมรรถนะของแบบจำลองต่าง ๆ ของข้อมูล NSL-KDD] [2] 15

ภาพที่ 7 เปรียบเทียบสมรรถนะของแบบจำลองต่าง ๆ ทดสอบด้วยข้อมูล AWID [2] 15

ภาพที่ 8 อัตราส่วนจำนวนข้อมูลสูญหายของตัวแปรทั้งหมด 17

ภาพที่ 9 อัตราส่วนของลูกค้ายกเลิกบริการและไม่ยกเลิกบริการของข้อมูลชุดที่หนึ่ง 20

ภาพที่ 10 อัตราส่วนของลูกค้ายกเลิกบริการและไม่ยกเลิกบริการของข้อมูลชุดที่สอง 20

ภาพที่ 11 กระบวนการทำนายของแบบจำลองการเรียนรู้เสริมกำลังเชิงลึก DQN 21

ภาพที่ 12 การเปลี่ยนสถานะเมื่อมีการกระทำที่แตกต่างกัน โดยสถานะมีความสัมพันธ์กัน 23

ภาพที่ 13 การเปลี่ยนสถานะของข้อมูลที่มีผู้สอน ที่แต่ละสถานะไม่มีความสัมพันธ์กัน 24

ภาพที่ 14 กระบวนการทำนายของแบบจำลองการเรียนรู้เสริมกำลัง Policy gradient 25

ภาพที่ 15 (ซ้าย) ค่าความแม่นยำเฉลี่ยของแบบจำลอง DQN ในแต่ละรอบการเรียนรู้ของข้อมูลสอนชุดที่หนึ่ง (ขวา) ค่าความแม่นยำเฉลี่ยของข้อมูลทดสอบชุดที่หนึ่ง 33

ภาพที่ 16 (ซ้าย) ค่าความแม่นยำเฉลี่ยของแบบจำลอง DQN ในแต่ละรอบการเรียนรู้ของข้อมูลสอนชุดที่สอง (ขวา) ค่าความแม่นยำเฉลี่ยของข้อมูลทดสอบชุดที่สอง 34

ภาพที่ 17 (ซ้าย) ค่าความแม่นยำเฉลี่ยของแบบจำลอง Policy gradient ในแต่ละรอบการเรียนรู้ของข้อมูลสอนชุดที่หนึ่ง (ขวา) ค่าความแม่นยำเฉลี่ยของข้อมูลทดสอบชุดที่หนึ่ง 34

ภาพที่ 18 (ซ้าย) ค่าความแม่นยำเฉลี่ยของแบบจำลอง Policy gradient ในแต่ละรอบการเรียนรู้ของข้อมูลสอนชุดที่สอง (ขวา) ค่าความแม่นยำเฉลี่ยของข้อมูลทดสอบชุดที่สอง 34

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

หัวใจสำคัญของธุรกิจหลายประเภท ได้แก่ ธุรกิจค้าปลีก ธุรกิจประกันภัย ธุรกิจโทรคมนาคม หรือ ธุรกิจการเงินการธนาคาร คือ ลูกค้าที่สร้างรายได้หลักให้กับธุรกิจเหล่านี้ นอกจากการหาลูกค้าใหม่แล้ว การรักษาฐานลูกค้าเดิมให้คงอยู่กับธุรกิจอย่างยาวนาน ถือเป็นสิ่งสำคัญอย่างมาก ทั้งนี้ การหาลูกค้าใหม่มีต้นทุนมากกว่า 5 เท่าของการรักษาลูกค้าเดิม [1] ในปัจจุบัน บริษัทต่าง ๆ จึงให้ความสำคัญกับการพัฒนาประสิทธิผลการทำนายพฤติกรรมลูกค้าที่มีแนวโน้มจะยกเลิกการใช้บริการ เนื่องจากการตรวจจับพฤติกรรมลูกค้าที่กำลังจะเปลี่ยนไปได้เร็ว และแม่นยำ จะทำให้บริษัทสามารถติดต่อลูกค้าเหล่านั้นได้โดยตรงอย่างรวดเร็ว เพื่อเสนอส่วนลดสินค้า หรือบริการพิเศษ เพื่อจูงใจให้ลูกค้าใช้บริการต่อไป ทำให้ช่วยลดความเสี่ยงจากการยกเลิกได้ และยังช่วยลดต้นทุนการสูญเสียลูกค้าเก่าได้ด้วย

ปัจจุบันมีงานวิจัยการทำนายพฤติกรรมลูกค้าที่จะยกเลิกบริการจำนวนมาก โดยจะใช้การเรียนรู้ของเครื่อง และการทำเหมืองข้อมูลในการทำนายว่าลูกค้าจะยกเลิกบริการหรือไม่ โดยส่วนมากแล้วลูกค้าที่ยกเลิกบริการ จะมีจำนวนน้อยกว่าลูกค้าที่ไม่ยกเลิกบริการ เป็นปัญหาข้อมูลไม่สมดุล (unbalanced data) หรือข้อมูลคำตอบของแต่ละตัวแปรป้ายมีจำนวนไม่เท่ากัน ในงานวิจัยส่วนใหญ่มีการจัดการกับข้อมูลที่ไม่สมดุลหลายวิธี โดยการสุ่มลดจำนวนข้อมูลกลุ่มหลักให้มีจำนวนข้อมูลใกล้เคียงกับกลุ่มน้อย (under-sampling) หรือการสุ่มเพิ่มจำนวนข้อมูลกลุ่มน้อยให้มีจำนวนใกล้เคียงกับกลุ่มหลัก (over-sampling) หรือวิธีการพิจารณาจากค่าความผิดพลาดจากการแบ่งกลุ่ม (cost-sensitive methods) นอกจากประเด็นข้อมูลไม่สมดุล ยังสังเกตได้ว่า แพตเทิร์นข้อมูลของลูกค้าอาจเปลี่ยนแปลงไปจากเดิมในอดีต จากสถานการณ์หนึ่งๆที่อาจส่งผลกระทบให้ค่าตัวแปรเปลี่ยนไป

การเรียนรู้เสริมกำลัง (Reinforcement Learning) เป็นการเรียนรู้ของเครื่องประเภทหนึ่ง ซึ่งมีความแตกต่างกับการเรียนรู้มีการชี้นำ (Supervised Learning) กล่าวคือ การเรียนรู้มีการชี้นำเป็นการเรียนรู้ที่ไม่โต้ตอบ (passive learning) นั่นคือ จะสามารถทำนายได้ดีกับข้อมูลที่ถูกสอนเท่านั้น ในขณะที่การเรียนรู้เสริมกำลังเป็นการเรียนรู้เชิงรุก (active learning) โดยระบบจะเรียนรู้จากการสื่อสารกับสิ่งแวดล้อม ซึ่งจะมีการเปลี่ยนแปลงไปตามการสำรวจแต่ละครั้ง ดังนั้น เมื่อมีข้อมูลใหม่ที่มีลักษณะที่แตกต่างกับข้อมูลเดิม หรือมีลักษณะการแจกแจงของข้อมูลที่แตกต่างไป การเรียนรู้มีการ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ชี้นำจะไม่สามารถทำนายได้ ต้องทำการเรียนรู้จากข้อมูลใหม่ทั้งหมด แต่การเรียนรู้เสริมกำลังจะปรับตัวกับสิ่งแวดล้อมนั้น ๆ ได้ทันที ดังนั้น การเรียนรู้เสริมกำลังจึงเป็นวิธีที่เหมาะกับปัญหาในโลกความเป็นจริงมากกว่าการเรียนรู้ของเครื่อง

Lopez-Martin และคณะ [2] ได้ประยุกต์ใช้การเรียนรู้เสริมกำลังเชิงลึก (Deep Reinforcement Learning) มาใช้ในการตรวจจับการโจมตี (Intrusion Detection) ที่เป็นโจทย์ของการเรียนรู้มีการชี้นำ กล่าวคือ ข้อมูลที่นำมาใช้ในการเรียนรู้มีตัวแปรป้าย (label) ผลการวิจัยพบว่าการเรียนรู้เสริมกำลังเชิงลึกสามารถจัดการกับข้อมูลที่ไม่สมดุลได้ดี โดยไม่จำเป็นต้องมีขั้นตอนการประมวลผลข้อมูลก่อน (data preprocessing) นอกจากนี้ ยังสามารถจัดการกับข้อมูลที่มีการแจกแจงของตัวแปรป้ายที่แตกต่างกันระหว่างชุดข้อมูลสอนและชุดข้อมูลทดสอบ การเรียนรู้เสริมกำลังเชิงลึกสามารถเพิ่มสมรรถนะการตรวจจับการโจมตีให้รวดเร็วและทำนายได้แม่นยำมากขึ้น เมื่อเปรียบเทียบผลลัพธ์กับแบบจำลองการเรียนรู้ของเครื่องอื่น ๆ ที่นิยมใช้กันในปัจจุบัน เช่น ซัพพอร์ตเวกเตอร์แมชีน (Support Vector Machine) การถดถอยโลจิสติกส์ (Logistic Regression) และวิธีเพื่อนบ้านใกล้ที่สุด k ตัว (K-Nearest Neighbors) ดังนั้น ผู้วิจัยจึงมีแนวคิดพัฒนาแบบจำลองด้วยวิธีการเรียนรู้เสริมกำลังเพื่อทำนายลักษณะลูกค้าที่จะยกเลิกบริการ เนื่องจากแบบจำลองสามารถปรับตัวได้เมื่อแพตเทิร์นในชุดข้อมูลมีการเปลี่ยนแปลงไปจากเดิม

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาสมรรถนะของการใช้เทคนิคการเรียนรู้เสริมกำลังในโจทย์การทำนายแบบมีผู้สอน โดยมีการวิเคราะห์ ออกแบบ และพัฒนาแบบจำลองเพื่อประเมินผลเปรียบเทียบกับแบบจำลองการเรียนรู้มีการชี้นำ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

1.3 ขอบเขตการดำเนินงาน

1. พัฒนาแบบจำลองการเรียนรู้เสริมกำลัง 2 แบบจำลอง คือ DQN และ Policy gradient
2. เปรียบเทียบสมรรถนะของแบบจำลองการเรียนรู้เสริมกำลังที่พัฒนาขึ้นกับแบบจำลองการเรียนรู้มีการชี้นำ ได้แก่ แรนดอมฟอเรสต์ เอกซ์จีบูสต์ และ kNN ตัว
3. ใช้ข้อมูลการจำแนกประเภท (Classification) ของลูกค้าในธุรกิจโทรคมนาคมแห่งหนึ่ง ซึ่งเป็นข้อมูลการใช้งานของลูกค้า ไม่รวมข้อมูลความคิดเห็นจากช่องทางอื่น เช่น ความคิดเห็นของลูกค้าต่อการบริการ

4. ประเมินผลด้วยค่าความแม่นยำ (accuracy) ค่าความเที่ยงตรง (precision) ค่าความครบถ้วน (recall) และค่าเอฟวัน (F1)

1.4 ขั้นตอนการวิจัย

1. ศึกษางานวิจัยเกี่ยวกับการนำการเรียนรู้เสริมกำลังมาใช้ในงานทำนายแบบมีผู้สอน
2. ศึกษางานวิจัยเกี่ยวกับการทำนายการยกเลิกบริการของลูกค้า
3. ศึกษาเทคนิคและซอฟต์แวร์การเรียนรู้เสริมกำลัง
4. รวบรวมข้อมูล
5. พัฒนาแบบจำลองการทำนายโดยใช้เทคนิคการเรียนรู้เสริมกำลัง และแบบจำลองการเรียนรู้มีการชี้นำ
6. ทดสอบและประเมินสมรรถนะของแบบจำลอง
7. วิเคราะห์ผลการทดลอง
8. เผยแพร่ผลงานวิชาการ
9. สรุปผลและเรียบเรียงวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

ได้แนวทางวิธีการและต้นแบบสำหรับการทำนายลักษณะลูกค้าที่จะยกเลิกบริการ และสามารถประยุกต์ใช้วิธีการที่นำเสนอกับปัญหาอื่นที่คล้ายกัน ที่ต้องการแบบจำลองที่สามารถปรับตัวได้ในกรณีที่ลักษณะหรือแพตเทิร์นข้อมูลมีความเป็นไปได้ที่จะเปลี่ยนแปลงไปจากเดิม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.6 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์

เนื้อหาในวิทยานิพนธ์แบ่งออกเป็น 6 บท ได้แก่ บทที่ 1 บทนำ อธิบายถึงที่มาและความสำคัญของปัญหา วัตถุประสงค์ของงานวิจัย ขอบเขตงานวิจัย ประโยชน์ที่คาดว่าจะได้รับ และผลงานที่ได้รับการตีพิมพ์จากวิทยานิพนธ์ บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง บทที่ 3 แนวคิดและวิธีการวิจัย บทที่ 4 การวิเคราะห์และประเมินผล และบทที่ 5 สรุปผลการวิจัย รวมถึงงานวิจัยในอนาคต

1.7 ผลงานที่ได้รับการตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์ฉบับนี้ได้รับการตีพิมพ์บทความทางวิชาการ

Panjasuchat, M., & Limpiyakorn, Y. (2020). Applying Reinforcement Learning for Customer Churn Prediction. *In 13th International Conference on Computer and Electrical Engineering*. Beijing, China.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 กระบวนการตัดสินใจแบบมาร์คอฟ (Markov Decision Process)

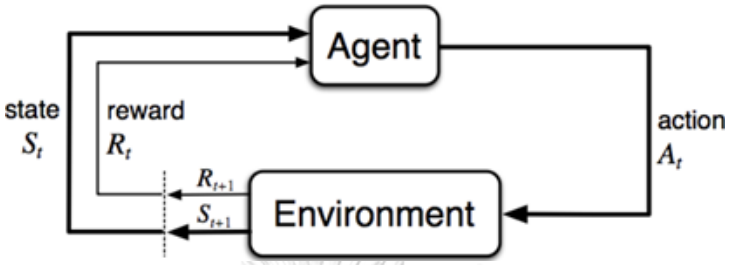
การเรียนรู้เสริมกำลังเป็นการเรียนรู้ที่มีพื้นฐานมาจากทฤษฎีกระบวนการตัดสินใจแบบมาร์คอฟ [3] โดยมีตัวแปรสำคัญ 4 ตัวแปร คือ เซตของสถานะ (state) แทนด้วย S เซตของการกระทำ (action) แทนด้วย A แผนผังที่กำหนดความน่าจะเป็นของการเปลี่ยนแปลงของคู่สถานะและการกระทำปัจจุบันไปเป็นสถานะใหม่ (transition) แทนด้วย T และผลรางวัล (reward) แทนด้วย R

สิ่งแวดล้อมที่มีคุณสมบัติแบบมาร์คอฟ (Markov property) คือ การที่สิ่งแวดล้อมมีการเปลี่ยนสถานะ โดยที่สถานะในอนาคตไม่ขึ้นกับสถานะในอดีต แต่จะขึ้นอยู่กับสถานะปัจจุบันเท่านั้น ทำให้การวิเคราะห์การเปลี่ยนแปลงสถานะของสิ่งแวดล้อมทำได้ง่ายขึ้น เนื่องจากไม่จำเป็นต้องรู้ถึงการเปลี่ยนแปลงสถานะในอดีตที่ผ่านมาทั้งหมด ถ้าหากสิ่งแวดล้อมมีคุณสมบัติแบบมาร์คอฟ จะเรียกสิ่งแวดล้อมนั้นว่ามีคุณสมบัติของกระบวนการตัดสินใจแบบมาร์คอฟ และหากสิ่งแวดล้อมนั้นมีจำนวนสถานะที่มีขอบเขต และมีจำนวนของการกระทำที่มีขอบเขต จะเรียกกระบวนการตัดสินใจนี้ว่ากระบวนการตัดสินใจแบบมาร์คอฟที่มีขอบเขตจำกัด (finite Markov Decision Process หรือ finite MDP) โดยในงานวิจัยนี้ จะใช้กระบวนการตัดสินใจแบบมาร์คอฟที่มีขอบเขตจำกัด เนื่องจากกำหนดให้ จำนวนสถานะ คือ จำนวนรูปแบบความเป็นไปได้ของข้อมูลนำเข้า นั่นก็คือจำนวนแถวของข้อมูล ซึ่งมีค่าเป็นจำนวนนับได้ และ การกระทำ คือ ยกเลิกบริการ หรือ ไม่ยกเลิกบริการ ก็มีค่าเป็นจำนวนนับได้ นั่นคือ มีขอบเขตจำกัด

2.1.2 การเรียนรู้เสริมกำลัง

การเรียนรู้เสริมกำลัง [3] คือ วิธีการเรียนรู้ที่เกิดจากการปฏิสัมพันธ์ระหว่าง ระบบ (Agent) กับสิ่งแวดล้อม (Environment) โดยตรง ระบบจะเรียนรู้จากค่าผลรางวัลที่ได้ เพื่อที่จะสามารถตัดสินใจเลือกการกระทำอย่างใดอย่างหนึ่งกับสิ่งแวดล้อม โดยมีเป้าหมายคือให้ได้ผลรางวัลรวมมากที่สุด สามารถเขียนเป็นวงรอบการทำงาน ดังภาพที่ 1 ซึ่งแสดงการทำงานโดยเริ่มจาก ระบบได้รับสถานะ และผลรางวัลที่เกิดขึ้น แล้วจึงทำการตัดสินใจเลือกการกระทำอย่างใดอย่างหนึ่งต่อสิ่งแวดล้อม จากนั้นสิ่งแวดล้อมจะให้ผลลัพธ์เป็นผลรางวัล เพื่อบอกว่าการกระทำนั้นดีหรือไม่ จากนั้น

สถานะเปลี่ยนไปเป็นสถานะใหม่ โดยระบบจะดำเนินไป และได้รับผลรางวัลกลับมาอยู่ตลอด จนกระทั่งระบบเกิดการเรียนรู้ว่า เมื่ออยู่ในสถานะใด แล้วจะต้องเลือกการกระทำใด ที่จะให้ผลรางวัลระยะยาว (long-term expected reward) สูงที่สุด โดยสามารถเขียนลำดับการทำงานได้ในรูปแบบ ..S_t, A_t, R_{t+1}, S_{t+1}, A_{t+1}, R_{t+2}, S_{t+2}, ...



ภาพที่ 1 การสื่อสารระหว่างระบบ และสิ่งแวดล้อมในกระบวนการตัดสินใจแบบมาร์คอฟ [3]

นิยามของส่วนประกอบต่าง ๆ ในวงรอบการทำงาน

agent คือ ตัวเรียนรู้และตัดสินใจภายใต้สถานการณ์ที่แตกต่างกันในระบบ

state S_t คือ สถานะของระบบ ณ เวลา t

reward r_t คือ ผลรางวัลในระยะยาวที่ได้จากการเลือกการกระทำ ณ เวลา t

action A_t คือ การกระทำที่ผู้เรียนรู้ตัดสินใจเลือกกระทำ ณ เวลา t

state S_{t+1} คือ สถานะของสิ่งแวดล้อม ณ เวลา t+1

reward R_{t+1} คือ ผลรางวัลที่ได้จากการกระทำ ณ เวลา t+1

เมื่อ s เป็นสถานะของสิ่งแวดล้อม S เป็นเซตสถานะทั้งหมดที่เป็นไปได้ โดย $s_t \in S$ และ $A(s_t)$ เป็นเซตของการกระทำทั้งหมดที่เป็นไปได้ในแต่ละสถานะ โดย $a_t \in A(s_t)$ และ $r_{t+1} \in \Re$ เมื่อ t คือเวลาในแต่ละขั้น โดย t = 0,1,2,3,...

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

องค์ประกอบของการเรียนรู้เสริมกำลัง ประกอบด้วย 4 องค์ประกอบหลัก คือ

ก) นโยบาย (policy) คือ ตัวกำหนดแนวทางการเรียนรู้สำหรับเลือกการกระทำในแต่ละสถานะ เพื่อให้ได้ผลรวมรางวัลระยะยาวเฉลี่ยสูงสุด ถ้าให้นโยบาย แทนด้วย π ที่ได้มาจากการหาค่าสูงสุดของผลเฉลี่ยรางวัลสะสมคิวแวลู แทนด้วย Q(s,a) เพื่อนำไปใช้เป็นฟังก์ชันในการเลือกการกระทำ ดังสมการ (1)

$$\pi(s) = \underset{a}{argmax}\, Q(s,a) \quad (1)$$

ข) สัญญาณผลรางวัล (reward signal) คือ ฟังก์ชันที่คำนวณผลรางวัลในระยะยาว โดยผลรางวัลนี้เกิดจากการเลือกการกระทำเมื่อระบบอยู่ในสถานะต่างๆ หากการกระทำนั้น

เป็นการตัดสินใจที่ถูกต้อง สภาพแวดล้อมจะให้ผลรางวัลในระดับสูง แต่หากเป็นการตัดสินใจที่ผิด จะให้ผลรางวัลในระดับที่ต่ำลงมา ซึ่งผลรางวัลนี้จะเป็นค่าที่แสดงถึงความสามารถในการตัดสินใจของระบบว่า สามารถทำการตัดสินใจได้ดีหรือไม่ในช่วงเวลานั้น หากพิจารณาผลรางวัลที่เกิดขึ้นภายหลังช่วงเวลา t กำหนดเป็น

$$g(s_t, a_t) + g(s_{t+1}, a_{t+1}) + g(s_{t+2}, a_{t+2}) + \cdots$$

โดยให้ผลรวมของผลรางวัลในระยะยาวแทนด้วย R_t จะได้ค่าผลรวมของผลรางวัลที่เกิดขึ้นตั้งแต่เวลา t จนถึง T ดังสมการ (2) ทั้งนี้ วิธีการเรียนรู้เสริมกำลังมีเป้าหมายที่ต้องการให้ R_t มีค่าสูงที่สุด

$$R_t = g(s_t, a_t) + g(s_{t+1}, a_{t+1}) + g(s_{t+2}, a_{t+2}) + \cdots + g(s_{T-1}, a_{T-1}) \quad (2)$$

ค) ฟังก์ชันมูลค่า (value function) คือ ฟังก์ชันที่ใช้ในการคำนวณหาผลรางวัลระยะยาวที่คาดว่าจะได้รับจากการกระทำนั้น ๆ ภายใต้นโยบายเดียวกัน ในการจำลองการกระทำ ซึ่งตัวกระทำการตัดสินใจ จะใช้ผลรางวัลนี้ในการตัดสินใจเลือกการกระทำในลำดับถัดไป โดยถ้าให้นโยบาย π เป็นนโยบายที่ทำการเชื่อมโยงระหว่างสถานะ s โดย $s \in S$ และการกระทำ a โดย $a \in A(s)$ และดำเนินไปภายใต้นโยบาย π นี้ ดังสมการ (3) ซึ่งเป็นสมการ สเตท-ฟังก์ชันมูลค่า สำหรับนโยบาย π

$$V^{\pi}(s) = E_{\pi}\{R_t | s_t = s\} = E_{\pi}\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s\} \quad (3)$$

เมื่อ $E_{\pi}\{\ \}$ คือ ค่าคาดหวังเมื่อตัวกระทำการตัดสินใจกระทำตามนโยบาย π และ γ เป็นอัตราการลดทอนที่มีค่าระหว่าง $0 \leq \gamma \leq 1$

ง) แบบจำลองของสภาพแวดล้อม (model of environment) คือ แบบจำลองของสภาพแวดล้อมที่จะนำการเรียนรู้เสริมกำลังไปประยุกต์ใช้ โดยต้องมีความสามารถที่จะแสดงพฤติกรรมได้เหมือนกับสภาพแวดล้อมจริงที่เราจะนำไปประยุกต์ใช้ ซึ่งต้องมีความสามารถที่จะแสดงพฤติกรรมได้เหมือนกับสภาวะแวดล้อมจริงที่จะนำไปประยุกต์ใช้ โดยถ้าให้ $P_{ss'}^{a}$ เป็นความน่าจะเป็นที่จะเกิดการเปลี่ยนสถานะจากสถานะ s เป็น s' เมื่อเลือกการกระทำ a ดังสมการที่ (4)

$$P_{ss'}^{a} = \Pr\{s_{t+1} = s' | s_t = s, a_t = a\} \quad (4)$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การนำวิธีการเรียนรู้เสริมกำลังมาใช้ในการแก้ปัญหานั้น จะเป็นการหานโยบายที่ทำให้เกิดผลรวมรางวัลระยะยาวสูงที่สุด โดยจะทำการปรับปรุงนโยบายให้ดีขึ้นอยู่ตลอด ดังนั้น นโยบายใหม่จึงดีกว่าหรือเท่ากับนโยบายเดิม ซึ่งทำให้ได้นโยบายอย่างน้อยหนึ่งนโยบายที่ดีกว่า หรือเท่ากับนโยบายอื่น ๆ นั่นก็คือ นโยบายที่เหมาะสมที่สุด (optimal policy) และเมื่อระบบดำเนินตามนโยบายนี้ จะได้สเตทฟังก์ชันมูลค่าที่เป็น สเตทแวลูที่เหมาะสม (optimal state-value function) ซึ่งแทนด้วย V* ดังสมการ (5) เมื่อทุก ๆ s ∈ S

$$V^*(s) = \max_{\pi} V^{\pi}(s) \tag{5}$$

เมื่อให้ระบบดำเนินไปตามนโยบายที่เหมาะสมนี้ จะได้แอ็กชันแวลูที่เป็น แอ็กชันแวลูที่เหมาะสม (optimal action-value function) ดังสมการ (6) เมื่อทุก ๆ s ∈ S และ a ∈ A(s)

$$Q^*(s,a) = \max_{\pi} Q^{\pi}(s,a) \tag{6}$$

สำหรับในทุก ๆ คู่ของสถานะและการกระทำ (s,a) จะได้ฟังก์ชันของค่าคาดหวังที่จะได้รับ เมื่อเลือกการกระทำ a และ สิ่งแวดล้อมอยู่ในสถานะ s และดำเนินไปด้วยนโยบายที่เหมาะสม จะได้ความสัมพันธ์ของแอ็กชันแวลูที่เหมาะสม Q* และ สเตทฟังก์ชันมูลค่าที่เหมาะสม V* ดังสมการ (7)

$$Q^*(s,a) = E\{r_{t+1}\gamma V^*(s_{t+1}) | s_t = s, a_t = a\} \tag{7}$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การหานโยบายที่เหมาะสม (optimal policy) [3] มี 2 วิธี กล่าวคือ

ก) วิธีมอนติคาร์โล (Monte Carlo Methods) เป็นวิธีการปรับค่าคิวของแต่ละคู่สถานะและการกระทำ โดยระบบจะทำการสุ่มการกระทำจากสถานะเริ่มต้น จนถึงสถานะสุดท้าย โดยจะทำการปรับค่าคิวของแต่ละคู่สถานะและการกระทำ หลังจากจบการทดลองในแต่ละครั้ง (episode) ดังสมการที่ 8 โดยวิธีมอนติคาร์โล มีข้อเสียที่ต้องรอจนกว่าจะจบการเรียนรู้ในแต่ละรอบ แล้วจึงจะทำการปรับค่าคิวของแต่ละคู่สถานะและการกระทำ ซึ่งต้องใช้เวลา และเปลืองทรัพยากรมากในการที่จะได้นโยบายที่เหมาะสมที่สุด

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(R_t - Q(S_t, A_t)) \tag{8}$$

ข) วิธี Temporal Difference Learning เป็นวิธีการปรับค่าคิวของแต่ละคู่สถานะและการกระทำ โดยทำการปรับค่าคิวทันทีหลังจากที่ระบบได้ตัดสินใจทำการกระทำในสถานะปัจจุบันจบและเปลี่ยนเป็นสถานะถัดไป ซึ่งแบบจำลองพื้นฐานจะมี 2 แบบจำลอง คือ

1) SARSA learning จะใช้นโยบายในการเลือกการกระทำของสถานะถัดไป แล้วจึงนำค่าคิวของสถานะถัดไปมาใช้ในการปรับค่าคิวของสถานะปัจจุบัน ดังสมการที่ 9

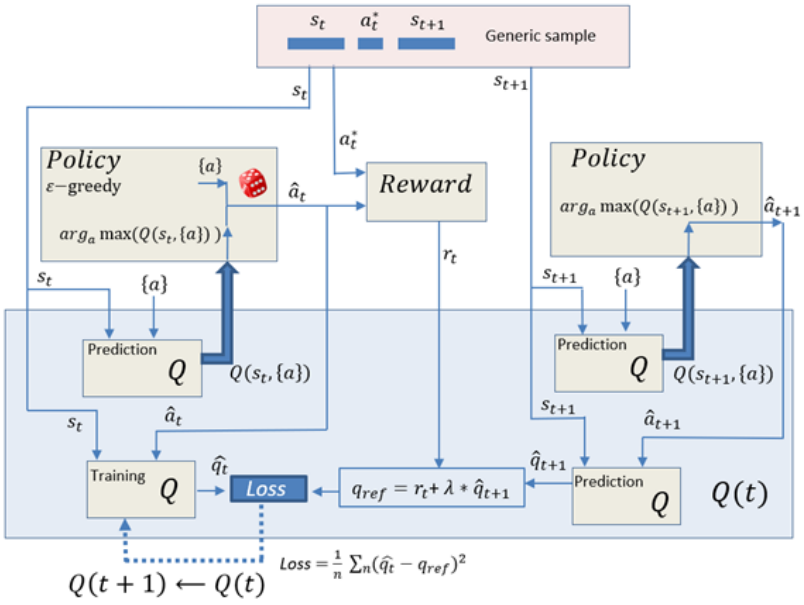
$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t)) \quad (9)$$

2) Q-learning จะทำการปรับค่าคิวของสถานะปัจจุบัน โดยใช้ค่าคิวที่มากที่สุดของสถานะถัดไป โดยไม่สนใจการกระทำของสถานะถัดไป ดังสมการที่ 10

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(R_{t+1} + \gamma max_{a \in A} Q(S_{t+1}, a) - Q(S_t, A_t)) \quad (10)$$

2.1.3 แบบจำลอง Deep Q Network (DQN)

แบบจำลอง DQN [2] เป็นการหานโยบายที่เหมาะสมโดยวิธี Temporal Different Learning แบบ Q-learning ภาพที่ 2 แสดงขั้นตอนการเรียนรู้ของ DQN ด้วยโครงข่ายประสาทเทียมเพื่อใช้ประมาณค่าคิวของสถานะและการกระทำปัจจุบัน โดยโครงข่ายประสาทเทียมจะประกอบด้วย 3 ชั้น คือ ชั้นข้อมูลนำเข้า (input layer) ชั้นซ่อน (hidden layer) และชั้นข้อมูลนำออก (output layer) และใช้ฟังก์ชันกระตุ้น ReLu ในทุกชั้น โดยชั้นข้อมูลนำเข้าจะมีจำนวนเท่ากับสถานะปัจจุบันและชั้นข้อมูลนำออกจะมีจำนวนเท่ากับการกระทำที่เป็นไปได้ ข้อมูลที่นำมาใช้ในการเรียนรู้ จะถูกแบ่งออกเป็นกลุ่มย่อย (batch) และประกอบด้วย 3 ค่า คือ 1) สถานะของระบบ ณ เวลา t 2) การกระทำระบบตัดสินใจเลือกกระทำ ณ เวลา t และ 3) สถานะของระบบ ณ เวลา t+1 โดยข้อมูลแต่ละแถวจะถูกส่งมาเรียนรู้โดยโครงข่ายประสาทเทียม เพื่อทำนายค่าคิวของสถานะและการกระทำ ณ เวลา t จากนั้น จะมีการนำค่าผิดพลาดค่าเฉลี่ยกำลังสอง (Mean Square Error Loss) ของค่าคิวที่ได้จากโครงข่ายประสาทเทียมสำหรับสถานะ ณ เวลา t และ ค่าคิวอ้างอิง ที่คำนวณมาจากผลรวมของค่าผลรางวัล ณ เวลา t และ ค่าคิวที่มากที่สุดของสถานะถัดไป คูณกับ ค่าปัจจัยคิดลด (discount factor) มาใช้ในการปรับน้ำหนักของโครงข่ายประสาทเทียม โดยจะทำการวนซ้ำข้อมูลจนครบทุกแถวในกลุ่มย่อยนี้ แล้วเปลี่ยนไปใช้ข้อมูลกลุ่มย่อยใหม่ จนครบทุกกลุ่มย่อย

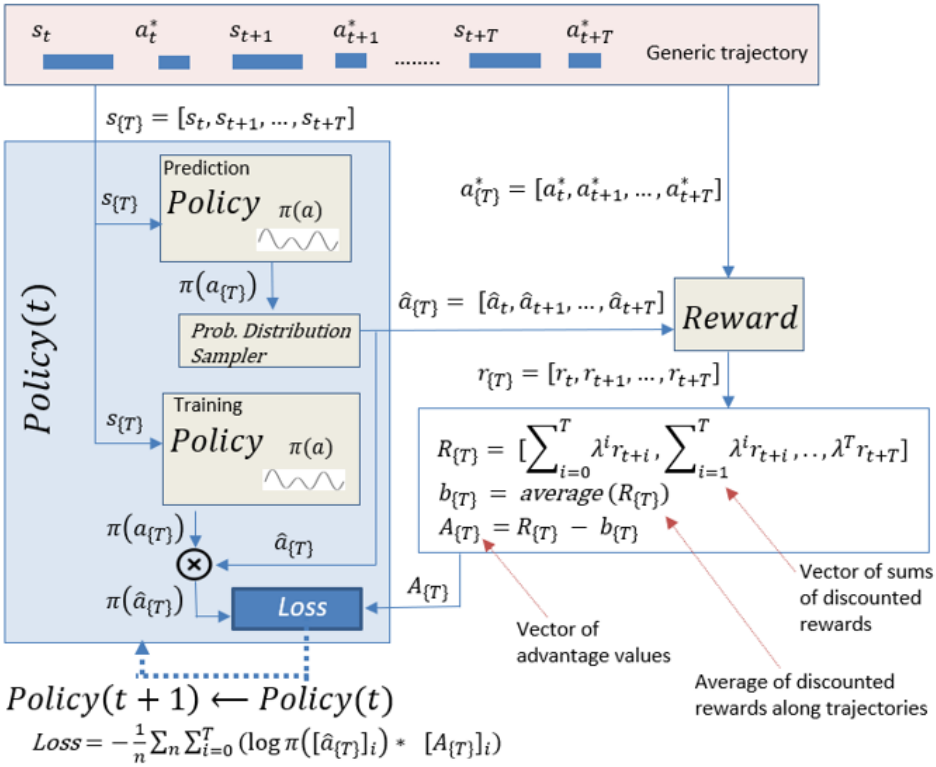


ภาพที่ 2 ขั้นตอนวิธีการเรียนรู้ของแบบจำลอง DQN [2]

2.1.4 แบบจำลอง Policy gradient

แบบจำลอง Policy gradient [2] เป็นการใช้โครงข่ายประสาทเทียม 1 โครงข่าย เพื่อทำนายนโยบายโดยตรง ภาพที่ 3 แสดงสถาปัตยกรรมแบบจำลอง Policy gradient ประกอบด้วยโครงข่ายประสาทเทียมซึ่งมีสามชั้น และใช้ฟังก์ชันกระตุ้นรีลู (ReLu) ในทุกชั้น แต่จะใช้ฟังก์ชันซอฟต์แมกซ์ (softmax) ในชั้นสุดท้าย เพื่อให้ได้ความน่าจะเป็นของการกระทำ ($\pi(a)$) โดยจะเริ่มจากการรับข้อมูลสถานะและตัวแปรป้ายแบบที่เป็นลำดับเข้ามา และนำสถานะไปเรียนรู้ที่โครงข่ายประสาทเทียม โดยให้ผลลัพธ์เป็นค่าความน่าจะเป็น แล้วระบบจะทำการสุ่มเลือกการกระทำจากค่าความน่าจะเป็นนั้น จากนั้นสิ่งแวดล้อมจะนำการกระทำของแต่ละสถานะไปเทียบกับตัวแปรป้ายแล้วส่งค่าผลรางวัลกลับมาให้ระบบ โดยผลรางวัลในที่นี้จะมีค่า 1 เมื่อระบบทำนายได้ถูกต้อง และมีค่า 0 เมื่อระบบทำนายไม่ถูกต้อง จากนั้นค่าผลรางวัลของแต่ละสถานะที่ได้รับ จะนำมาคูณกับปัจจัยคิดลด แล้วนำมาหาค่าเฉลี่ยเพื่อเป็นค่าเบสไลน์ ขั้นตอนต่อไป ระบบจะคำนวณหาผลต่างของผลรางวัลแต่ละสถานะและค่าเบสไลน์ได้เป็นค่าประโยชน์ (advantage values) เพื่อนำค่านี้ไปใช้ในการปรับค่าน้ำหนักของโครงข่าย สำหรับฟังก์ชันที่ใช้หาค่าผิดพลาด จะใช้ฟังก์ชันล็อก-ลอส (log-loss function) สำหรับกระบวนการทำนาย ซึ่งเลือกการกระทำที่ให้ค่าความน่าจะเป็นสูงที่สุด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 3 ขั้นตอนวิธีการเรียนรู้ของแบบจำลอง Policy gradient [2]

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 งานวิจัยเกี่ยวกับการทำนายลักษณะลูกค้าที่จะยกเลิกบริการ

Rai, S. และคณะ [4] นำเสนอการสร้างแบบจำลองเพื่อทำนายการยกเลิกบริการโดยใช้ต้นไม้ตัดสินใจ ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลขนาดเล็ก จำนวน 20,000 แถว และจำนวนตัวแปร 12 ตัวแปร โดยไม่มี data preprocessing ผู้วิจัยใช้ 10-fold cross validation ในการทดสอบสมรรถนะของแบบจำลอง พบว่าต้นไม้ตัดสินใจให้ค่าความแม่นยำมากกว่าการวิเคราะห์ถดถอยโลจิสติกส์ มีค่าเท่ากับ 0.7025 และ 0.6403 ตามลำดับ

Keramati, A.และคณะ [5] ใช้เทคนิคการทำเหมืองข้อมูลทั้งหมด 4 แบบจำลอง คือ ต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม ขั้นตอนวิธีเพื่อนบ้านใกล้เคียง k ตัว และซัพพอร์ตเวกเตอร์แมชีน หลังจากนั้นได้เสนอวิธีไฮบริด โดยการรวมทั้ง 4 ระบบมาใช้ในการทำนายการยกเลิกบริการ โดยผลการวิจัยพบว่า ใน 4 แบบจำลอง โครงข่ายประสาทเทียมให้ค่า recall มากที่สุด แต่วิธีไฮบริดให้ค่า precision และค่า recall มากที่สุดเมื่อเทียบกับเมื่อใช้แบบจำลองทั้ง 4 แบบเดี่ยว

นอกจากนี้ ปัญหาข้อมูลไม่สมดุลดังที่กล่าวข้างต้น เป็นปัญหาหลักอย่างหนึ่งในงานการทำนายลักษณะลูกค้าที่ยกเลิกบริการ โดยงานวิจัยต่าง ๆ มีการจัดการกับปัญหานี้หลายวิธี

งานวิจัยของ Cao, S. และคณะ [6] ได้นำเสนอโครงข่ายประสาทเทียม stacked autoencoder โดยเริ่มต้นการฝึกแบบจำลองโดยใช้การเรียนรู้แบบไม่มีการชี้นำแบบ greedy layerwise ในการฝึกสอนแต่ละชั้น หลังจากผ่านการฝึกสอนแต่ละชั้นแล้ว จะนำมารวมกันเป็นสแต็กและมีการใช้ขั้นตอนวิธีส่งค่าย้อนกลับ (backpropagation algorithm) เพื่อปรับค่าพารามิเตอร์ให้แม่นยำขึ้น หลังจากนั้นจะใช้การวิเคราะห์การถดถอยโลจิสติกส์ในการจัดกลุ่มลูกค้าว่าจะยกเลิกหรือไม่ยกเลิกบริการ โดยวิธีนี้ช่วยแก้ปัญหาในการจัดการกับข้อมูลขนาดใหญ่ และข้อมูลที่ตัวแปรป้ายเป็นเป้าหมายมีจำนวนน้อย นอกจากนี้ยังช่วยลดมิติของข้อมูลได้

ผู้วิจัยได้แก้ปัญหาข้อมูลไม่สมดุล โดยการสุ่มลดจำนวนข้อมูลกลุ่มหลักให้มีจำนวนใกล้เคียงกับกลุ่มน้อย ซึ่งในงานวิจัยนี้มีข้อมูลทั้งหมด 300,000 แถว เป็นลูกค้าที่ยกเลิกบริการ 100,000 คน จึงทำการสุ่มข้อมูลกลุ่มลูกค้าที่ไม่ยกเลิกบริการให้เหลือ 100,000 คน แล้วนำข้อมูลทั้งหมด 200,000 คนไปทำการฝึกสอนแบบจำลอง จากการทดลอง ผู้วิจัยพบว่ายิ่งจำนวนข้อมูลลดลงจะทำให้ค่าความแม่นยำ (accuracy) มีค่าเพิ่มขึ้น สุดท้ายผู้วิจัยใช้ข้อมูลจำนวน 12,000 แถวในการฝึกสอน ให้ค่าความแม่นยำสูงสุดที่ 73.5% ซึ่งมากกว่าแบบจำลองการวิเคราะห์ถดถอยโลจิสติกส์เพียงเล็กน้อย

งานวิจัยของ Ahmad, A. และคณะ [7] ได้พัฒนาแบบจำลองการทำนายลักษณะลูกค้าที่จะยกเลิกบริการ โดยใช้เทคนิคการเรียนรู้ของเครื่องบนแพลตฟอร์มของข้อมูลขนาดใหญ่ นอกจากนี้ยังมีการนำตัวแปรที่เกี่ยวกับการวิเคราะห์เครือข่ายทางสังคมมาใช้การทำนายด้วย โดยแบบจำลองที่ใช้ประกอบด้วย Decision Tree, Random Forest, Gradient Boosted Machine และ XGBoost โดยผลการวิจัยพบว่า XGBoost ให้ค่าความแม่นยำมากที่สุด และการเพิ่มตัวแปรที่เกี่ยวกับการวิเคราะห์เครือข่ายทางสังคม ส่งผลให้ความแม่นยำในการทำนายเพิ่มมากขึ้นอีกด้วย

ข้อมูลที่นำมาใช้ในการวิจัยนี้ เป็นข้อมูลขนาดใหญ่และไม่สมดุล ผู้วิจัยจึงได้ทำการทดลองกับการจัดการข้อมูลที่ไม่สมดุล โดยแบ่งเป็น 3 การทดลอง คือ 1. การสุ่มเพิ่มจำนวนข้อมูลกลุ่มน้อยให้มีจำนวนใกล้เคียงกับข้อมูลกลุ่มหลัก (Oversampling) 2. การสุ่มลดจำนวนข้อมูลกลุ่มหลักให้มีจำนวนใกล้เคียงกับข้อมูลกลุ่มน้อย (Undersampling) และ 3. ไม่ทำการสุ่มข้อมูลเพื่อให้สมดุล (without re-balancing) จากผลการวิจัย พบว่า แบบจำลอง XGBoost และ Gradient Boosted Machine ให้ผลลัพธ์ดีสุดกับข้อมูลที่ไม่ทำการสุ่มข้อมูลเพื่อให้สมดุล แต่แบบจำลอง Random

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

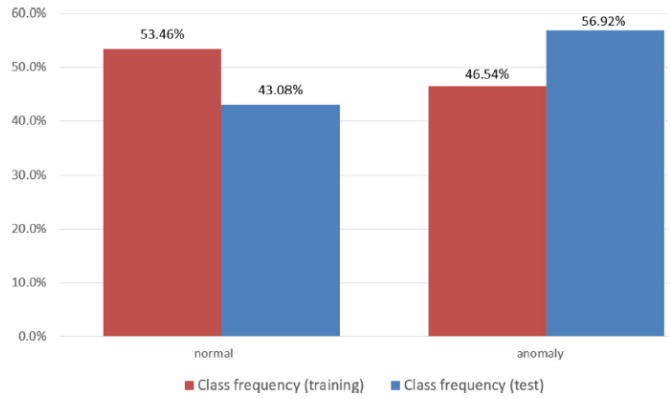
Forest และ ต้นไม้ตัดสินใจ ให้ผลลัพธ์ดีที่สุดกับข้อมูลที่ทำการสุ่มลดจำนวนข้อมูลกลุ่มหลักให้มีจำนวนใกล้เคียงกับข้อมูลกลุ่มน้อย

Wong, M.L. และคณะ [8] ได้เสนอวิธีการจัดการกับข้อมูลที่ไม่สมดุล 2 วิธี คือ Cost-Sensitive Deep neural network (CSDNN) ซึ่งเป็นการพิจารณาค่าความผิดพลาดจากการแบ่งกลุ่มโดยใช้โครงข่ายประสาทเทียมเชิงลึกแบบ Stacked Denoising Autoencoders และ Cost-Sensitive Deep neural network ensemble (CSDE) โดยการรวมการเรียนรู้ของวิธี CSDNN ใช้การสุ่มลดจำนวนข้อมูลกลุ่มน้อย และทำการสกัดคุณสมบัติทีละชั้นของชั้นซ่อน (layer-wise feature extraction) โดยทำการทดลองกับชุดข้อมูลจริง 6 ชุดจากหลากหลายธุรกิจ และมีข้อมูลพฤติกรรมการยกเลิกของลูกค้าเป็นหนึ่งในชุดข้อมูลที่ใช้ทดลอง ในการประเมินสมรรถนะของแบบจำลอง ผู้วิจัยได้ทำการเปรียบเทียบกับแบบจำลองอื่น ๆ อีกมากมายทั้งAdaCost, MetaCost, RUSBoost, RBBoost, SMOTEBoost, AdaBoost, LogiBoost, BalanceCascade, Bagging, Voting, Logistic regression, Neural network, Support vector machine, Bayesian network and Decision tree ผลการวิจัยพบว่าวิธี CSDE ให้ผลลัพธ์ที่ดีที่สุด

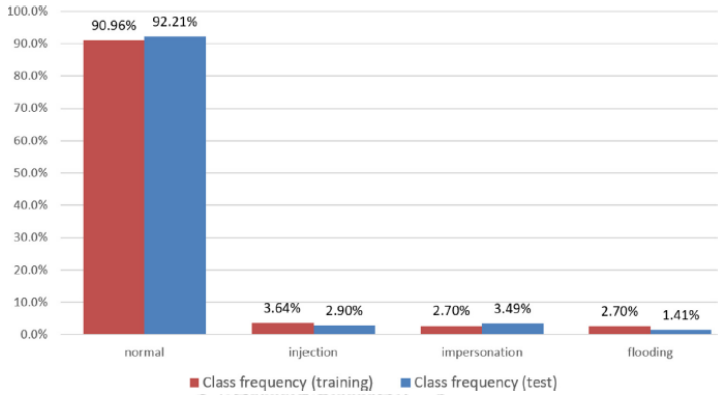
2.2.2 งานวิจัยเกี่ยวกับการใช้เทคนิคการเรียนรู้เสริมกำลังกับโจทย์การเรียนรู้มีการชี้นำ

Lopez-Martin, M. และคณะ [2] ได้นำเสนอการใช้อัลกอริทึมการเรียนรู้เสริมกำลังเชิงลึก (Deep Reinforcement Learning) ในระบบการตรวจจับการโจมตี (Intrusion Detection System) โดยแบบจำลองที่ใช้ในงานวิจัยนี้มีทั้งหมด 4 แบบ คือ Deep Q-Network (DQN), Double Deep Q-Network (DDQN), Policy Gradient (PG) และ Actor-Critic (AC) ผู้วิจัยทำการทดลองกับชุดข้อมูลที่เป็นที่นิยมในงานวิจัยตรวจจับการโจมตี ทั้งหมด 2 ชุด คือ NSL-KDD และ AWID โดยข้อมูลมีขนาดใหญ่ และค่อนข้างทันสมัย แต่ทั้งสองชุดเป็นข้อมูลที่ไม่สมดุล โดย NSL-KDD มีการแจกแจงของตัวแปรป้าย (label) ของชุดข้อมูลสอน (Training set) และ ชุดข้อมูลทดสอบ (Test set) แตกต่างกัน ดังภาพที่ 4 AWID มีตัวแปรป้ายมากกว่าสองค่า และ ไม่สมดุลอย่างมาก โดยที่ 90 % ของตัวแปรป้ายมีค่าเดียวกัน ดังภาพที่ 5

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 4 อัตราส่วนของตัวแปรป้ายระหว่างข้อมูลสอนและข้อมูลทดสอบของ NSL-KDD [2]



ภาพที่ 5 อัตราส่วนของตัวแปรป้ายระหว่างข้อมูลสอนและข้อมูลทดสอบของ AWID [2]

การประเมินสมรรถนะของแบบจำลอง จะใช้ข้อมูลชุดทดสอบกับแบบจำลองในการเรียนรู้เสริมกำลัง 4 แบบ: DQN, DDQN, Policy gradient และ Actor Critic เปรียบเทียบกับแบบจำลองการเรียนรู้ของเครื่องอื่นๆ ประกอบด้วย ซัพพอร์ตเวกเตอร์แมชีน การวิเคราะห์ถดถอยโลจิสติกส์ ตัวจำแนกเบส์อย่างง่าย เพื่อนบ้านใกล้ที่สุด k ตัว แรนดอมฟอเรสต์ เกรเดียนต์บูสติงแมชีน เอดาบูสต์ เพอร์เซปตรอนหลายชั้น และ โครงข่ายประสาทเทียมแบบคอนโวลูชัน โดยตัววัดที่ใช้ในการประเมินผลของแต่ละแบบจำลอง ประกอบด้วย ค่าความแม่นยำ (accuracy) ค่าเอฟวัน (F1) ค่าความเที่ยงตรง (precision) และ ค่าความครบถ้วน (recall) จากการทดลองกับข้อมูลชุด NSL-KDD ได้ผลลัพธ์ว่า DDQN, SVM RBF-kernel, DQN, Actor critic และ Naïve Bayes แบบ Bernoulli, all features, quantizing the continuous features ให้ค่าความแม่นยำมากที่สุด เรียงตามลำดับ ดังภาพที่ 6 และในเรื่องของเวลาที่ใช้ในการเรียนรู้ และในการทำนาย พบว่า แบบจำลองห้าอันดับที่มีค่าความแม่นยำสูงที่สุดตามที่ได้กล่าวมาข้างต้น แบบจำลอง Naïve Bayes ใช้เวลาในการเรียนรู้เร็วที่สุด (64 วินาที) รองลงมาคือ DQN และ DDQN (290 และ 500 วินาที) SVM และ Actor critic ใช้เวลาในการเรียนรู้ค่อนข้างมาก (ประมาณ 1,700 วินาที) แต่เวลาที่ใช้ในการทำนายผล พบว่า แบบจำลอง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

DQN, DDQN และ Actor critic ใช้เวลาไม่ถึง 1 วินาที (0.6 วินาที) รองลงมาคือ Naïve Bayes (1.12 วินาที) และสุดท้าย SVM (159 วินาที) ดังภาพที่ 6

		Test results			
		Accuracy	F1	Precision	Recall
Logistic Regression	Logistic Regression	0.7068	0.6807	0.8955	0.5491
SVM	SVM, Linear Kernel	0.7740	0.7718	0.9073	0.6715
	SVM, RBF Kernel	0.8799	0.8927	0.9081	0.8779
KNN	K-Nearest Neighbors	0.7808	0.7769	0.9233	0.6706
Random Forest	Random Forest	0.7472	0.7211	0.9688	0.5743
GBM	Gradient Tree Boosting	0.7761	0.7612	0.9690	0.6287
Naive Bayes	Gaussian, only continous features	0.5490	0.7018	0.5627	0.9324
	Bernoulli, only discrete features	0.7942	0.7979	0.9046	0.7138
	Bernoulli, all features, quantizing the continous features	0.8019	0.7967	0.9583	0.6818
AdaBoost	Adaboost with Trees	0.7806	0.7403	0.9683	0.5992
	Adaboost with Naive Bayes	0.5265	0.3181	0.8824	0.1940
Neural Network	Neural Net (MLP)	0.7966	0.7881	0.9679	0.6647
CNN	CNN-1D	0.7875	0.7633	0.8094	0.7875
Reinforcement Learning	DQN	0.8787	0.8935	0.8833	0.8937
	DDQN	0.8978	0.9120	0.8944	0.9303
	Policy Gradient	0.7873	0.7909	0.8980	0.7067
	Actor Critic	0.8078	0.8111	0.9203	0.7251

		Training time (sec)	Prediction time (sec)
Logistic Regression	Multinomial	97.37	0.55
SVM	SVM, Linear Kernel	65.06	0.46
	SVM, RBF Kernel	1696.16	158.65
KNN	KNN	91.53	80.30
Random Forest	Random Forest	97.31	3.87
GBM	Gradient Tree Boosting	2242.14	4.39
Naive Bayes	Gaussian, only continous features	3.04	0.23
	Bernoulli, only discrete features	5.52	0.36
	Bernoulli, quantized features	63.89	1.12
AdaBoost	Adaboost, Naive Bayes	1466.73	113.22
Neural Network	Neural Net (MLP)	314.74	0.89
CNN	CNN-1D	590.58	1.52
Reinforcement Learning	DQN	290.50	0.54
	DDQN	507.01	0.55
	Policy Gradient	352.48	0.65
	Actor Critic	1725.17	0.83

ภาพที่ 6 ผลการทดลองสมรรถนะของแบบจำลองต่าง ๆ ของข้อมูล NSL-KDD] [2]

ผลการทดลองของข้อมูลชุด AWID ได้ผลลัพธ์ว่า ต้นไม้ตัดสินใจ J48, แรนดอมฟอเรสต์, DDQN, DQN และ MLP เป็น 5 แบบจำลองที่มีค่าความแม่นยำมากที่สุดตามลำดับ รวมถึงค่าเอฟวัน ค่าความถูกต้อง และค่าความครบถ้วน ดังภาพที่ 7 จึงสรุปได้ว่า แบบจำลองที่ใช้การเรียนรู้เสริมกำลังเชิงลึกให้สมรรถนะการทำนายที่ดี และเป็นแบบจำลองที่แข็งแกร่ง เนื่องจากให้ผลลัพธ์ที่ดีกับทุกเหตุการณ์ ทั้งข้อมูลที่มีอัตราส่วนของตัวแปรป้ายไม่เหมือนกัน และข้อมูลที่ไม่สมดุล

		Test results			
		Accuracy	F1	Precision	Recall
Boosting	AdaBoost	0.9220	0.8850	0.8500	0.9220
Decision Tree	J48	0.9620	0.9480	0.9620	0.9630
Frequency Table based	Hyper pipes	0.9223	0.8850	0.8790	0.9220
	Naive Bayes	0.9055	0.9090	0.9170	0.9060
	ZeroR	0.9220	0.8850	0.8500	0.9220
	OneR	0.9457	0.9220	0.9000	0.9460
Random Forest	Random Forest	0.9582	0.9440	0.9590	0.9580
Neural Network	MLP	0.9470	0.9256	0.9174	0.9473
Reinforcement Learning	DQN	0.9541	0.9372	0.9244	0.9541
	DDQN	0.9570	0.9394	0.9235	0.9570
	Policy Gradient	0.9221	0.8847	0.8502	0.9221
	Actor Critic	0.9221	0.8847	0.8502	0.9221

ภาพที่ 7 เปรียบเทียบสมรรถนะของแบบจำลองต่าง ๆ ทดสอบด้วยข้อมูล AWID [2]

บทที่ 3
แนวคิดและวิธีวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาสมรรถนะของแบบจำลองการเรียนรู้เสริมกำลังในการประยุกต์ใช้ทำนายการยกเลิกบริการของลูกค้า เมื่อแพตเทิร์นของข้อมูลมีการเปลี่ยนแปลงไปจากเดิม โดยในบทนี้จะกล่าวถึงข้อมูลที่นำมาใช้ การเตรียมข้อมูล โครงสร้างของแบบจำลองที่ใช้ในงานวิจัย และการวัดสมรรถนะแบบจำลองตามแนวทางที่นำเสนอ

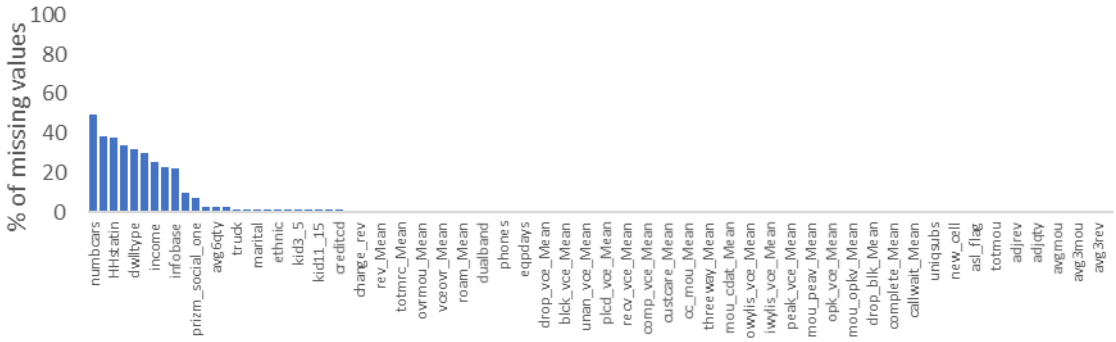
3.1 ข้อมูลที่ใช้ในงานวิจัยและการประมวลผลข้อมูลเบื้องต้น

ข้อมูลที่นำมาใช้ในงานวิจัยนี้ เป็นข้อมูลสาธารณะที่ได้จาก www.kaggle.com ซึ่งเกี่ยวกับข้อมูลการใช้งานและประวัติลูกค้าของบริษัทโทรคมนาคมแห่งหนึ่ง ซึ่งมีคะแนนของการนำข้อมูลไปใช้งานอยู่ที่ 7.1 เป็นคะแนนที่ทาง Kaggle พิจารณาจากข้อมูลนั้นว่าสามารถใช้งานได้ง่าย และครบถ้วนสมบูรณ์หรือไม่ [9] โดยข้อมูลแต่ละแถวคือข้อมูลของลูกค้าแต่ละคน ประกอบด้วยข้อมูล 100,000 แถว และตัวแปรทั้งหมด 100 ตัวแปร แบ่งเป็นตัวแปรเชิงคุณภาพ 21 ตัวแปร โดย 1 ตัวแปรเป็นไอดีของลูกค้า และ 20 ตัวแปรเกี่ยวกับข้อมูลส่วนตัวของลูกค้า เช่น เพศ สถานภาพ เป็นต้น ตัวแปรเชิงปริมาณ 79 ตัวแปรที่เกี่ยวกับสถิติการโทรศัพท์ของลูกค้า เช่น จำนวนนาทีเฉลี่ยของการโทรศัพท์ในช่วง 6 เดือนที่ผ่านมา หรือร้อยละการเปลี่ยนแปลงของจำนวนนาทีเฉลี่ยต่อเดือนเทียบกับค่าเฉลี่ยของสามเดือนที่ผ่านมา เป็นต้น โดยขั้นตอนในการประมวลผลข้อมูลเบื้องต้นมีดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3.1.1 การทำความสะอาดข้อมูล (data cleaning)

จากการสำรวจข้อมูล พบว่ามีตัวแปร 9 ตัวแปรที่มีอัตราส่วนจำนวนข้อมูลสูญหายมากกว่า 20% ของจำนวนข้อมูลทั้งหมด ดังภาพที่ 8 นอกจากนี้ยังมีตัวแปรอีก 2 ตัวแปรที่ไม่สมบูรณ์เนื่องจากความผิดพลาดจากกระบวนการเก็บข้อมูล คือ วงเงินของบัญชี (asl_flag) และ ดัชนีเครดิตการ์ด (creditcd) จึงทำการตัดตัวแปรทั้งหมด 12 ตัวแปร ประกอบด้วย 11 ตัวแปรข้างต้น และ 1 ตัวแปรที่เป็นไอดีของลูกค้าออกจากการวิเคราะห์ หลังจากนั้นจึงพิจารณาตัดแถวข้อมูลที่มีข้อมูลสูญหายอย่างน้อยหนึ่งค่า ทำให้จำนวนแถวของข้อมูลทั้งหมดจะเท่ากับ 78,334 แถวและ 88 ตัวแปร



ภาพที่ 8 อัตราส่วนจำนวนข้อมูลสูญหายของตัวแปรทั้งหมด

3.1.2 การแปลงข้อมูล (Data transformation)

ทำการแปลงตัวแปรเชิงคุณภาพทั้งหมด 14 ตัวแปรให้เป็นตัวแปรเชิงปริมาณ โดยการทำวันฮอท เอ็นโค้ดเดอร์ (one hot encoder) คือ การสร้างตัวแปรจากตัวแปรเชิงคุณภาพให้เป็นตัวแปรย่อย ๆ ที่มีค่าเป็นไปได้สองค่า (binary) นั่นคือ 0 หมายถึง ข้อมูลแถวนั้นไม่มีลักษณะตามตัวแปรนั้น และ 1 หมายถึง ข้อมูลแถวนั้นมีลักษณะตามตัวแปรนั้น

3.1.3 การคัดเลือกตัวแปร (Feature selection)

ทำการเลือกตัวแปรที่มีความสำคัญกับการทำนาย โดยพิจารณาจากค่าความสำคัญของตัวแปร (Feature importance) ที่ได้จากแบบจำลองแรนดอมฟอเรสต์ เพื่อลดขนาดของตัวแปรที่ใช้ในการสร้างแบบจำลอง คะแนนความสำคัญเป็นคะแนนที่บอกถึงความสามารถในการแบ่งกลุ่มข้อมูลตามตัวแปรป้ายได้อย่างถูกต้องของตัวแปรนั้น ๆ ยิ่งมีค่ามาก หมายถึงมีความสามารถในการแบ่งกลุ่มได้ดี จึงทำการคัดเลือกตัวแปรที่มีคะแนนสูงที่สุด 21 ตัวแปร ซึ่งตัวแปรทั้ง 21 ตัวนี้เป็นตัวแปรเชิงปริมาณที่บอกเกี่ยวกับข้อมูลการใช้งานของลูกค้า ดังตารางที่ 1

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 1 ชื่อตัวแปร ความหมาย และคะแนนความสำคัญของ 21 ตัวแปร

ชื่อตัวแปร	ความหมาย	คะแนนความสำคัญ
eqpdays	อายุปัจจุบันของอุปกรณ์	0.0309
change_mou	ร้อยละการเปลี่ยนแปลงของจำนวนนาทีเฉลี่ยต่อเดือนเทียบกับค่าเฉลี่ยสามเดือนที่ผ่านมา	0.0265
months	จำนวนเดือนที่ใช้บริการ	0.0217
mou_Mean	จำนวนนาทีเฉลี่ยที่ใช้ต่อเดือน	0.0210
adjrev	ค่าใช้จ่ายทั้งหมดตั้งแต่เริ่มใช้บริการ	0.0207
totrev	ค่าใช้จ่ายทั้งหมด	0.0212
change_rev	ร้อยละการเปลี่ยนแปลงของค่าใช้จ่ายเฉลี่ยต่อเดือนเทียบกับค่าเฉลี่ยสามเดือนที่ผ่านมา	0.0207
avgrev	ค่าใช้จ่ายเฉลี่ยต่อเดือนตลอดการใช้งาน	0.0204
avgmou	จำนวนนาทีที่ใช้เฉลี่ยต่อเดือน	0.0203
avgqty	จำนวนครั้งที่โทรเฉลี่ยต่อเดือน	0.0220
totcalls	จำนวนครั้งที่โทรทั้งหมด	0.0195
adjqty	จำนวนครั้งที่โทรทั้งหมดตั้งแต่เริ่มใช้บริการ	0.0194
totmou	จำนวนนาทีที่ใช้ทั้งหมด	0.0194
rev_Mean	ค่าใช้จ่ายเฉลี่ยต่อเดือน	0.0196
avg3mou	จำนวนนาทีที่ใช้เฉลี่ยในช่วงสามเดือนที่ผ่านมา	0.0192
adjmou	จำนวนนาทีที่ใช้ทั้งหมดตั้งแต่เริ่มใช้บริการ	0.0193
avg6mou	จำนวนนาทีที่ใช้เฉลี่ยในช่วงหกเดือนที่ผ่านมา	0.0189
avg3qty	จำนวนครั้งที่โทรเฉลี่ยในช่วงสามเดือนที่ผ่านมา	0.0186
avg6qty	จำนวนครั้งที่โทรเฉลี่ยในช่วงหกเดือนที่ผ่านมา	0.0178
mou_cvce_Mean	จำนวนนาทีเฉลี่ยที่การโทรสมบูรณ์	0.0177
mou_opkv_Mean	จำนวนนาทีเฉลี่ยที่ใช้โทรในช่วงเวลาที่ไม่ใช่ช่วงนิยม	0.0178

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.1.4 เตรียมชุดข้อมูลสำหรับการทดลองข้อมูลมีแพตเทิร์นเปลี่ยนไปจากเดิม

หลังจากขั้นตอน 3.1.1, 3.1.2 และ 3.1.3 ที่ได้กรองความสมบูรณ์ของตัวแปร แปลงข้อมูลและคัดเลือกตัวแปรที่สำคัญเรียบร้อยแล้ว จะได้ชุดข้อมูลที่มีจำนวนแถวทั้งหมด 78,334 แถว และ 22 ตัวแปร โดยข้อมูลทั้งหมดแบ่งเป็นลูกค้าที่ยกเลิกบริการ 37,974 แถวและลูกค้าที่ไม่ยกเลิกบริการ 40,360 แถว คิดเป็นอัตราส่วนของลูกค้าที่ยกเลิกและไม่ยกเลิกบริการ เท่ากับ 48% และ 52% ตามลำดับ จากนั้นจะทำการเตรียมชุดข้อมูลเพื่อใช้ในการทดลองข้อมูลมีแพตเทิร์นเปลี่ยนไปจากเดิม

แพตเทิร์นของข้อมูลที่เปลี่ยนแปลงไปในงานวิจัยนี้ หมายถึง ลูกค้ามีพฤติกรรมการใช้งานที่เปลี่ยนแปลงไปจากเดิม เช่น เมื่อในอดีตลูกค้าเป็นลูกค้าไม่ยกเลิกบริการ แต่ปัจจุบันลูกค้าเปลี่ยนสถานะเป็นยกเลิกบริการ ผู้วิจัยจึงจะทำการจำลองข้อมูลหนึ่งชุด ที่มีตัวแปรป้ายของลูกค้าเปลี่ยนไปจากเดิม โดยมีขั้นตอนดังต่อไปนี้

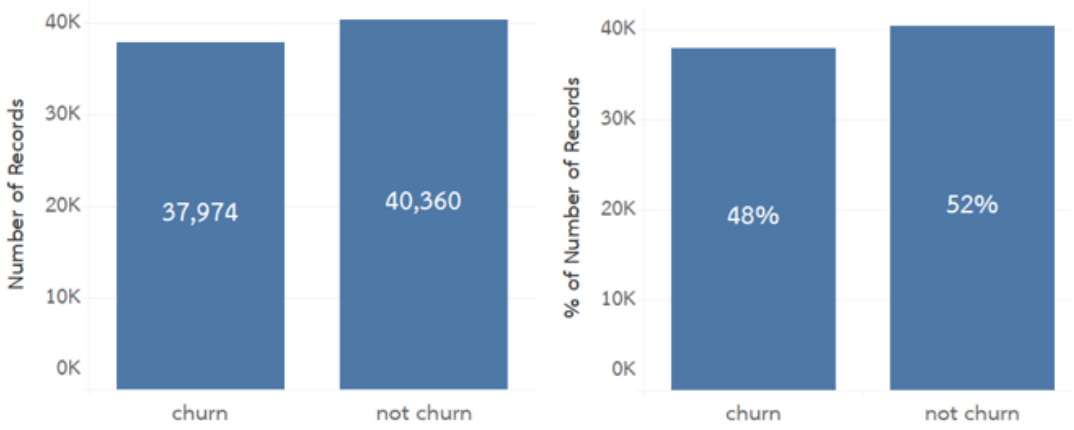
1. แบ่งข้อมูลตามคลาส จะได้ข้อมูลเป็นสองกลุ่ม คือ ลูกค้ายกเลิกบริการ และลูกค้าไม่ยกเลิกบริการ

2. ทำการสุ่มข้อมูลลูกค้ายกเลิกบริการมา 18,987 คน หรือ 50% ของจำนวนลูกค้ายกเลิกบริการทั้งหมด แล้วทำการเปลี่ยนตัวแปรป้ายจากยกเลิกบริการ เป็นไม่ยกเลิกบริการ

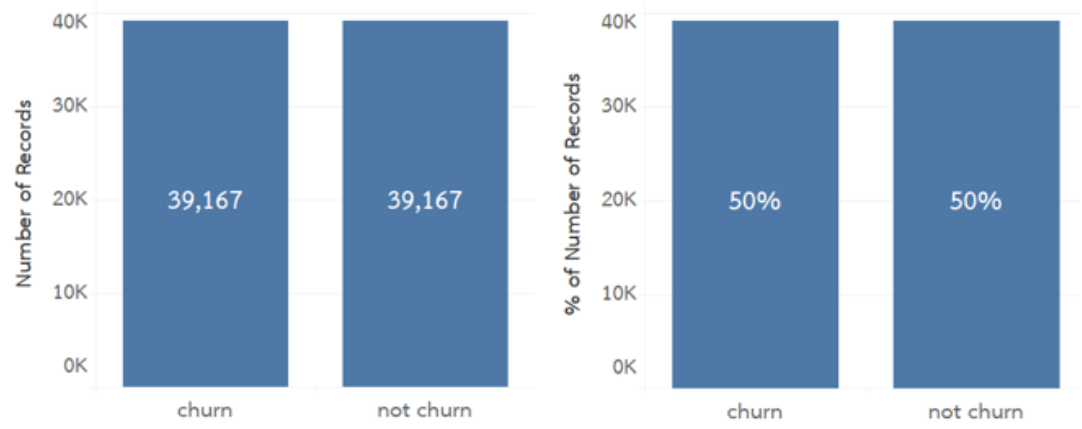
3. ทำการสุ่มข้อมูลลูกค้าไม่ยกเลิกบริการมา 20,180 คน หรือ 50% ของจำนวนลูกค้าไม่ยกเลิกบริการทั้งหมด แล้วทำการเปลี่ยนตัวแปรป้ายจากไม่ยกเลิกบริการ เป็นยกเลิกบริการ

4. รวมข้อมูลจากข้อ 2. และ 3. จะได้ว่าลูกค้ายกเลิกบริการมีทั้งหมด 39,167 คน โดย 18,987 คนเป็นลูกค้ายกเลิกบริการเดิม และอีก 20,180 คนเดิมเป็นลูกค้าไม่ยกเลิกบริการ แต่เปลี่ยนเป็นยกเลิกบริการตามข้อ 3. นอกจากนี้กลุ่มลูกค้าไม่ยกเลิกบริการจะมีทั้งหมด 39,167 คน โดย 20,180 คนเป็นลูกค้าไม่ยกเลิกบริการเดิม และอีก 18,987 คนเดิมเป็นลูกค้ายกเลิกบริการ แต่เปลี่ยนเป็นไม่ยกเลิกบริการตามข้อ 2

ภาพที่ 9 แสดงอัตราส่วนของลูกค้ายกเลิกบริการและไม่ยกเลิกบริการของข้อมูลชุดที่หนึ่ง พบว่าเป็นข้อมูลที่ค่อนข้างสมดุล โดยมีอัตราส่วนของลูกค้ายกเลิกบริการและไม่ยกเลิกบริการเท่ากับ 48%:52% และจากภาพที่ 10 จะเห็นว่าอัตราส่วนของลูกค้ายกเลิกบริการและไม่ยกเลิกบริการของข้อมูลชุดที่สองเท่ากับ 50%:50%



ภาพที่ 9 อัตราส่วนของลูกค้ายกเลิกบริการและไม่ยกเลิกบริการของข้อมูลชุดที่หนึ่ง



ภาพที่ 10 อัตราส่วนของลูกค้ายกเลิกบริการและไม่ยกเลิกบริการของข้อมูลชุดที่สอง

หลังจากได้ข้อมูลสองกลุ่มข้างต้นแล้ว จะมีการสร้างข้อมูลชุดที่สาม ซึ่งเป็นการรวมกันระหว่างข้อมูลชุดที่หนึ่ง และข้อมูลชุดที่สอง โดยข้อมูลชุดนี้จะใช้ในการสร้างแบบจำลองของการเรียนรู้แบบมีการชี้นำ นั่นก็คือ แรนดอมฟอเรสต์ เอกซ์จีบูสต์ และวิธีหาเพื่อนบ้านใกล้เคียง k ตัว จากนั้นแบ่งข้อมูลแต่ละกลุ่มออกเป็นสองชุด คือ ข้อมูลสอน และข้อมูลทดสอบด้วยอัตราส่วน 80%:20% จะได้จำนวนแถว และอัตราส่วนข้อมูลตัวแปรป้ายดังตารางที่ 2

ตารางที่ 2 อัตราส่วนของตัวแปรป้ายในชุดข้อมูลสอน และข้อมูลทดสอบกรณีแพตเทิร์นของข้อมูลเปลี่ยนแปลงไปจากเดิม

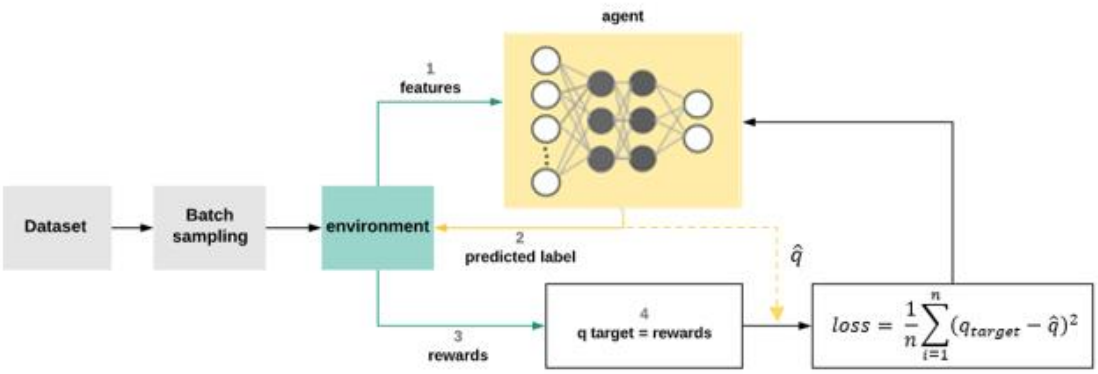
		จำนวนแถวข้อมูล			อัตราส่วนแถวข้อมูล		
		ยกเลิก	ไม่ยกเลิก	รวม	ยกเลิก	ไม่ยกเลิก	รวม
ข้อมูลชุดที่หนึ่ง	ข้อมูลสอน	30,379	32,288	62,667	48%	52%	100%
	ข้อมูลทดสอบ	7,595	8,072	15,667	48%	52%	100%
	รวม	37,974	40,360	78,334	48%	52%	100%
ข้อมูลชุดทีสอง (แพตเทิร์นเปลี่ยน)	ข้อมูลสอน	31,334	31,333	62,667	50%	50%	100%
	ข้อมูลทดสอบ	7,833	7,834	15,667	50%	50%	100%
	รวม	39,167	39,167	78,334	50%	50%	100%
ข้อมูลชุดที่สาม	ข้อมูลสอน	61,713	63,621	125,334	49%	51%	100%
	ข้อมูลทดสอบ	15,428	15,906	31,334	49%	51%	100%
	รวม	77,141	79,527	156,668	49%	51%	100%

3.2 โครงสร้างของแบบจำลองการเรียนรู้เสริมกำลัง

ในการใช้แบบจำลองการเรียนรู้เสริมกำลังในการทำนาย จะต้องแปลงข้อมูลให้อยู่ในกระบวนการตัดสินใจแบบมาร์คอฟ นั่นคือ สถานะ (state (s)) คือ เวกเตอร์ตัวแปรทั้งหมด และการกระทำ (action (a)) คือ ตัวแปรป้าย ในที่นี้เป็นได้สองค่า คือ ยกเลิกบริการ หรือ ไม่ยกเลิกบริการ งานวิจัยนี้จะใช้แบบจำลองการเรียนรู้เสริมกำลังสองแบบจำลอง คือ DQN และ Policy gradient

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.2.1 ขั้นตอนการเรียนรู้ของแบบจำลองการเรียนรู้เสริมกำลัง DQN



ภาพที่ 11 กระบวนการทำนายของแบบจำลองการเรียนรู้เสริมกำลังเชิงลึก DQN

แบบจำลองการเรียนรู้เสริมกำลัง DQN ประกอบด้วยส่วนที่เป็นระบบ (agent) และสิ่งแวดล้อม (environment) โดยระบบจะทำหน้าที่ในการเรียนรู้แต่ละสถานะ และทำการตัดสินใจเลือกการกระทำที่ให้ค่าผลรางวัลมากที่สุด โดยระบบในแบบจำลอง DQN จะเป็นโครงข่ายประสาทเทียมที่ประกอบด้วยชั้นทั้งหมด 4 ชั้น นั่นคือ 1) ชั้นนำเข้า (input layer) ที่มีจำนวนโหนดเท่ากับจำนวนตัวแปรต้น เท่ากับ 21 โหนด 2) ชั้นซ่อน (hidden layer) ทั้งหมด 2 ชั้นที่มีจำนวนโหนดเท่ากับ 512 โหนดในแต่ละชั้น และ 3) ชั้นนำออก (output layer) ที่มีจำนวนโหนดเท่ากับจำนวนการกระทำที่เป็นไปได้ ในที่นี้เท่ากับ 2 โหนด สำหรับฟังก์ชันกระตุ้นจะใช้ฟังก์ชันเรคคิไฟท์เชิงเส้น (ReLU function) ในทุกชั้น ยกเว้นชั้นนำออกจะใช้ฟังก์ชันเชิงเส้น (linear function) เนื่องจากต้องการให้ผลลัพธ์ที่ออกมาเป็นผลบวกของค่านำเข้าและค่าน้ำหนัก โดยไม่มีฟังก์ชันอื่นที่มาปรับค่าผลลัพธ์นั้นอีก สำหรับฟังก์ชันต้นทุน จะใช้ค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean Squared Error: MSE) ดังสมการที่ 11 และใช้อัลกอริทึม Adam ในการปรับปรุงค่าผิดพลาดเพื่อให้ได้ค่าที่ดีที่สุด

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{11}$$

ขั้นตอนการเรียนรู้ของแบบจำลองการเรียนรู้เสริมกำลัง DQN นั้นจะเรียนรู้ข้อมูลเป็นชุด (batch) โดยระบบจะทำการเรียนรู้ซ้ำหลาย ๆ รอบ (episode) ในแต่ละรอบจะมีการนำค่าผลรางวัลที่ได้รับจากสิ่งแวดล้อมมาปรับค่าน้ำหนักของโครงข่าย เพื่อให้ความสามารถในการทำนายมีความแม่นยำมากยิ่งขึ้น โดยขั้นตอนการเรียนรู้ของแบบจำลอง สามารถอธิบายได้ดังภาพที่ 11 เริ่มจากสุ่มข้อมูลเป็นชุดเล็ก ในงานวิจัยนี้จะสุ่มข้อมูลชุดละ 5,000 แถว และจะทำการเรียนรู้ทั้งหมด 200 รอบ ในแต่ละรอบมีการทำงานดังนี้ ระบบจะทำการเรียนรู้ โดยรับสถานะจากสิ่งแวดล้อม นั่นคือ อาเรย์ของตัวแปรต้นทั้งหมด 21 ตัว ดังนั้นข้อมูลที่ระบบรับเข้ามาในแต่ละรอบ จะเป็นอาเรย์ที่มีมิติ 5,000 แถว และ 21 คอลัมน์ จากนั้นระบบจะเรียนรู้โดยใช้โครงข่ายประสาทเทียม และในชั้นนำออกจะส่งค่าคิวที่ทำนายออกมาเป็นอาเรย์ที่มีมิติ 5,000 แถว และ 2 คอลัมน์ตามการกระทำที่เป็นไปได้ทั้งหมด นั่นคือ ยกเลิก หรือไม่ยกเลิก เนื่องจากข้อมูลในการวิเคราะห์นี้ การกระทำของแต่ละสถานะ จะไม่เกี่ยวข้องกับสถานะอื่น จะสิ้นสุดที่สถานะของตัวเองเท่านั้น ทำให้ระบบจะตัดสินใจเลือกการกระทำจากการกระทำที่ให้ค่าคิว ณ สถานะนั้นมีค่ามากที่สุด ดังสมการที่ 12

$$a = argmax_a(Q(s,a)) \tag{12}$$

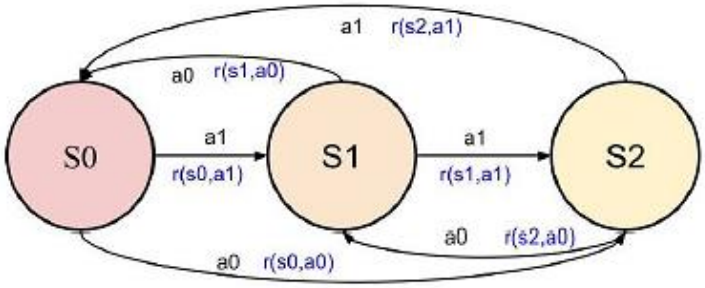
จากนั้นสิ่งแวดล้อมจะพิจารณาการกระทำกับตัวแปรป้ายของแต่ละสถานะว่าถูกต้องหรือไม่ แล้วส่งเป็นค่าผลรางวัล (reward) ที่เป็นอาเรย์ที่มีมิติ 5,000 แถว และ 1 คอลัมน์กลับมาที่ระบบเพื่อใช้ในการปรับค่าน้ำหนักของโครงข่ายประสาทเทียม โดยค่าผลรางวัลจะกำหนดเป็นฟังก์ชันว่า ถ้าการกระทำที่ระบบทำนายตรงกับตัวแปรป้าย จะให้ค่าผลรางวัลเท่ากับ 1 แต่ถ้าการกระทำที่ระบบทำนายไม่ตรงกับตัวแปรป้าย จะให้ค่าผลรางวัลเท่ากับ 0 จากนั้นจะเป็นขั้นตอนการเรียนรู้ของระบบ โดยระบบจะนำค่าผลรางวัลที่ได้นี้มาหาค่าคิวทาร์เก็ต ฟังก์ชันที่ใช้หาค่าคิวทาร์เก็ต โดยปกติแล้วจะเท่ากับผลรวมของผลรางวัล ณ สถานะนั้นและ ผลคูณของตัวปัจจัยคิดลด (discount factor) และค่าคิวที่มากที่สุดของสถานะถัดไป ดังสมการที่ 13

$$Q_{target} = r_{t+1} + \gamma max_a Q(s_{t+1}, a) \tag{13}$$

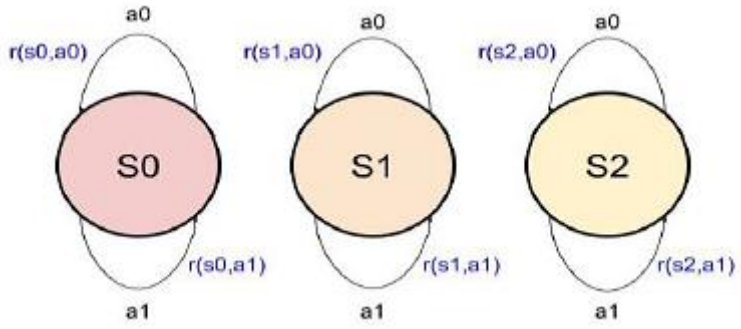
เนื่องจากแต่ละสถานะในข้อมูลชุดนี้ จะสิ้นสุดที่สถานะของตัวเองเท่านั้น ดังภาพที่ 13 ดังนั้น จึงสามารถตัดพจน์ที่เป็นค่าคิวที่มากที่สุดของสถานะถัดไปจากสมการได้ ค่าคิวทาร์เก็ตของแต่ละสถานะก็จะเท่ากับค่าผลรางวัลของสถานะนั้น ๆ ดังสมการที่ 14 โดยคิวทาร์เก็ตจะเป็นอาเรย์ที่มีมิติ 5,000 แถว และ 2 คอลัมน์ ถ้าการกระทำเป็นคลาส 0 จะให้คิวทาร์เก็ตของคอลัมน์คลาส 0 เท่ากับค่าผลรางวัลที่ได้ และคิวทาร์เก็ตของคอลัมน์คลาส 1 เท่ากับค่าติดลบของผลรางวัลที่ได้ ในทางกลับกัน ถ้าการกระทำเป็นคลาส 1 จะให้คิวทาร์เก็ตของคอลัมน์คลาส 1 เท่ากับค่าผลรางวัลที่ได้ และคิวทาร์เก็ตของคอลัมน์คลาส 0 เท่ากับค่าติดลบของผลรางวัลที่ได้ จากนั้นจะนำค่าคิวที่ได้จากโครงข่าย และคิวทาร์เก็ตที่ได้ข้างต้น ไปใช้ในการเรียนรู้ปรับค่าน้ำหนักของโครงข่ายประสาทเทียม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$Q_{target} = r_{t+1} \tag{14}$$



ภาพที่ 12 การเปลี่ยนสถานะเมื่อมีการกระทำที่แตกต่างกัน โดยสถานะมีความสัมพันธ์กัน



ภาพที่ 13 การเปลี่ยนสถานะของข้อมูลที่มีผู้สอน ที่แต่ละสถานะไม่มีความสัมพันธ์กัน

3.2.2 ขั้นตอนการเรียนรู้ของแบบจำลองการเรียนรู้เสริมกำลัง Policy gradient

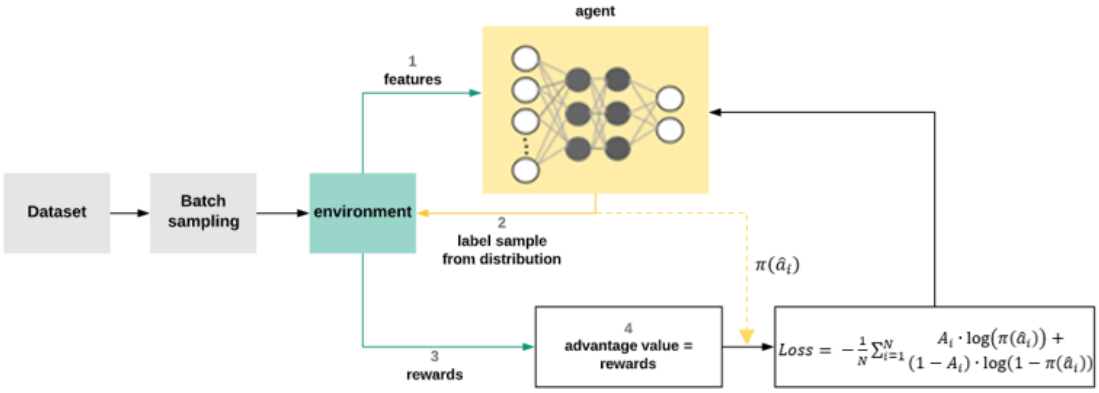
ขั้นตอนการเรียนรู้ของแบบจำลองการเรียนรู้เสริมกำลัง Policy gradient จะแตกต่างจากแบบจำลอง DQN ตรงที่ DQN ระบบจะเรียนรู้ค่าคิวของแต่ละสถานะและการกระทำ และนโยบาย (policy) ที่ใช้ในการเลือกการกระทำของแต่ละสถานะก็จะดูจากค่าคิวที่มากที่สุด แต่ Policy gradient ระบบจะทำการเรียนรู้นโยบายโดยตรง โดยระบบจะมีโครงข่ายประสาทเทียมหนึ่งโครงข่ายที่ไว้ทำนายค่าความน่าจะเป็นของแต่ละการกระทำ โครงข่ายประสาทเทียมนี้ประกอบด้วย 4 ชั้น นั่นคือ 1) ชั้นนำเข้า มีจำนวน 21 โหนดเท่ากับจำนวนตัวแปรต้น 2) ชั้นซ่อน ทั้งหมด 2 ชั้นที่มีจำนวนโหนดเท่ากับ 1,048 โหนดในแต่ละชั้น และ 3) ชั้นนำออก มีจำนวนโหนดเท่ากับ 2 โหนดตามจำนวนการกระทำที่เป็นไปได้ทั้งหมด สำหรับฟังก์ชันกระตุ้นจะใช้ฟังก์ชันเรคคิไฟท์เชิงเส้น (ReLU function) เป็นฟังก์ชันกระตุ้นในทุกชั้น ยกเว้นชั้นนำออกที่จะใช้ฟังก์ชันซอฟต์แม็กซ์ (Softmax function) เพื่อให้ผลลัพธ์ที่ออกมาเป็นค่าความน่าจะเป็น สำหรับฟังก์ชันต้นทุน จะใช้ไบนารี ครอส-เอนโทรปี (Binary Cross-Entropy) ดังสมการที่ 15 และใช้อัลกอริทึม Adam ในการปรับปรุงค่าผิดพลาดเพื่อให้ได้ค่าที่ดีที่สุด

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log\big(p(y_i)\big) + (1-y_i) \cdot \log(1-p(y_i)) \qquad (15)$$

ขั้นตอนการเรียนรู้ดังภาพที่ 14 เริ่มต้นจากระบบรับข้อมูลที่เป็นอาเรย์ที่มีมิติเท่ากับจำนวนแถวของข้อมูลทั้งหมด และจำนวนคอลัมน์เท่ากับ 21 ตามจำนวนตัวแปรต้น จากนั้นโครงข่ายประสาทเทียมจะให้ค่าผลลัพธ์ออกมาเป็นค่าความน่าจะเป็นของการกระทำแต่ละการกระทำ จะได้เป็นอาเรย์ที่มีมิติเท่ากับ จำนวนแถวของข้อมูล และจำนวนคอลัมน์เท่ากับ 2 ตามจำนวนการกระทำที่เป็นไปได้ทั้งหมด แล้วระบบจะสุ่มเลือกการกระทำจากความน่าจะเป็นที่ได้ จากนั้นสิ่งแวดล้อมจะส่ง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าผลรางวัลกลับมา ถ้าการกระทำที่ทำนายตรงกับตัวแปรป้าย จะให้ค่าเท่ากับ 1 แต่ถ้าการกระทำที่ทำนายไม่ตรงกับตัวแปรป้าย จะให้ค่าเท่ากับ -1 และเนื่องจากสถานะไม่มีความสัมพันธ์กันดังภาพที่ 14 จึงจะนำค่าผลรางวัลนี้มาใช้เป็นค่าประโยชน์ของสถานะนั้น ๆ เลย จากนั้นจะเป็นขั้นตอนการเรียนรู้ของระบบ โดยการค่าประโยชน์ของแต่ละสถานะของข้อมูลในรอบนั้น มาคำนวณค่าผิดพลาดไบนารี ครอส-เอนโทรปี ดังสมการที่ 15 แล้วใช้ในการปรับค่าน้ำหนักของโครงข่ายประสาทเทียม



ภาพที่ 14 กระบวนการทำนายของแบบจำลองการเรียนรู้เสริมกำลัง Policy gradient

3.3 แนวทางการวัดสมรรถนะของแบบจำลอง

ในการวัดสมรรถนะแบบจำลองในงานวิจัยนี้จะใช้ตัววัดการจำแนกแบบสองคลาส (Binary classification) โดยสามารถหาได้จากตารางคอนฟิวชันเมทริกซ์ ดังตารางที่ 3

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 3 คอนฟิวชันเมทริกซ์ที่ใช้จำแนกแบบสองคลาส

		ตัวแปรป้ายที่ทำนาย	
		ยกเลิก	**ไม่ยกเลิก**
ตัวแปรป้ายจริง	**ยกเลิก**	True Positive (TP)	False Negative (FN)
	ไม่ยกเลิก	False Positive (FP)	True Negative (TN)

ค่าที่ใช้วัดสมรรถนะการจำแนกมีทั้งหมด 4 ค่า ดังนี้

1. ค่าความแม่นยำ (accuracy) เป็นการวัดความแม่นยำในการทำนายโดยรวมของแบบจำลอง นั่นก็คือ แบบจำลองทำนายถูกกี่ครั้งจากจำนวนที่ทำนายทั้งหมด สามารถคำนวณได้จากสมการที่ 16

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (16)$$

2. ค่าความเที่ยงตรง (precision) เป็นการวัดแบบจำลองว่าสามารถทำนายคำตอบได้ถูกต้องเท่าไร โดยพิจารณาจากตัวแปรป้ายที่สนใจ สามารถคำนวณได้จากสมการที่ 17

$$Precision = \frac{TP}{TP+FP} \quad (17)$$

3. ค่าความครบถ้วน (recall) เป็นการวัดแบบจำลองว่าจากตัวแปรป้ายที่สนใจแบบจำลองสามารถทำนายค่าได้ถูกต้องครบถ้วนเท่าไร สามารถคำนวณได้จากสมการที่ 18

$$Recall = \frac{TN}{TN+FN} \quad (18)$$

4. ค่าเอฟวัน (F1) เป็นการวัดทั้งค่าความเที่ยงตรงและความครบถ้วน โดยจะเป็นค่าเฉลี่ยแบบฮาร์โมนิกของทั้งสองค่า สามารถคำนวณได้จากสมการที่ 19

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (19)$$

จากที่ได้กล่าวในข้อ 3.1.4 ข้อมูลที่ใช้ในเหตุการณ์แพตเทิร์นเปลี่ยนไปมีทั้งหมดสองชุด ดังนั้นในการวัดผลของแบบจำลอง DQN และ Policy gradient จะมีการวัดผลจากชุดข้อมูลสองชุด นั่นคือ ข้อมูลชุดทีหนึ่ง เป็นข้อมูลปกติ และข้อมูลชุดสอง เป็นข้อมูลที่มีแพตเทิร์นเปลี่ยนไป โดยแบบจำลองการเรียนรู้เสริมกำลัง DQN และ Policy gradient เป็นการเรียนรู้แบบออนไลน์ นั่นคือจะทำการเรียนรู้ไปพร้อมกับการทำนาย ดังนั้นในการวัดผลของแบบจำลอง จะเริ่มต้นจากการใช้ข้อมูลสอนชุดหนึ่งสร้างแบบจำลอง จากนั้นจะเริ่มขั้นตอนการวัดผล โดยการสุ่มข้อมูลรอบละ 5,000 แถว แล้วทำตามขั้นตอนการเรียนรู้ของแบบจำลองตามที่กล่าวไว้ในหัวข้อ 3.2.1 และ 3.2.2 โดยจะเรียนรู้ทั้งหมด 200 รอบ ในแต่ละรอบจะมีการบันทึกค่าวัดทั้งสี่ จนครบรอบที่ 200 ซึ่งเป็นรอบที่สามารถทำนายได้แม่นยำมากที่สุด จึงจะใช้ค่าที่ได้จากการวัดรอบที่ 200 นี้มาใช้ในการเปรียบเทียบกับแบบจำลองอื่น จากนั้นจะให้แบบจำลองได้เรียนรู้กับข้อมูลชุดสอง ข้อมูลที่มีแพตเทิร์นเปลี่ยนไป และจะทำการเรียนรู้และวัดผลเช่นเดียวกันกับข้อมูลชุดทีหนึ่ง แต่จะใช้ข้อมูลทดสอบของข้อมูลชุดสองในการวัดผลแทน และจะใช้ค่าวัดทั้งสี่ของรอบที่ 200 ในการเปรียบเทียบกับแบบจำลองอื่น

สำหรับการวัดผลของแบบจำลองการเรียนรู้แบบมีการชี้นำทั้งสาม คือ แรนดอมฟอเรสต์ เอกซ์จีบูสต์ และวิธีเพื่อนบ้านใกล้ที่สุด k ตัว เนื่องจากในหัวข้อ 3.1.4 จากตารางที่

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2 ในข้อมูลชุดทีหนึ่งและข้อมูลชุดสอง จะมีการแบ่งข้อมูลออกเป็นสองชุด คือ ข้อมูลสอน และข้อมูลทดสอบ แต่เนื่องจากต้องมีการปรับจูนค่าไฮเปอร์พารามิเตอร์ จึงจะต้องมีข้อมูลตรวจสอบที่ไว้ใช้ในการปรับจูน โดยจะเตรียมจากการแบ่งจากข้อมูลสอนด้วยอัตราส่วนของข้อมูลสอนและข้อมูลตรวจสอบเท่ากับ 70%:30% จากนั้นเริ่มต้นโดยการใช้ข้อมูลสอนของข้อมูลชุดทีหนึ่งในการสร้างแบบจำลอง แล้วปรับจูนค่าไฮเปอร์พารามิเตอร์โดยใช้ข้อมูลตรวจสอบ จากนั้นวัดผลของแบบจำลองที่หนึ่งโดยใช้ข้อมูลทดสอบของข้อมูลชุดทีหนึ่ง จากนั้นจะทำการสร้างแบบจำลองใหม่จากข้อมูลชุดสาม ซึ่งเป็นข้อมูลที่รวมทั้งข้อมูลปกติ และข้อมูลที่มีแพตเทิร์นเปลี่ยนไป เพื่อให้แบบจำลองสามารถทำนายกับข้อมูลที่เปลี่ยนไปได้ จากนั้นปรับจูนค่าไฮเปอร์พารามิเตอร์โดยใช้ข้อมูลตรวจสอบของข้อมูลชุดสาม และวัดผลของแบบจำลองที่สองนี้โดยใช้ข้อมูลทดสอบของข้อมูลชุดสาม

ดังนั้น ในงานวิจัยนี้จะมีการเปรียบเทียบสมรรถนะของแบบจำลองจากสองชุดข้อมูล คือ ชุดข้อมูลปกติ และชุดข้อมูลที่มีแพตเทิร์นเปลี่ยนไป เพื่อประเมินว่าแบบจำลองการเรียนรู้เสริมกำลังและแบบจำลองการเรียนรู้ที่มีการชี้นำ สามารถปรับตัวได้กับการที่มีข้อมูลที่เปลี่ยนแปลงไปหรือไม่ และยังมีความสามารถในการทำนายที่ดีหรือไม่จากค่าตัววัดทั้งสี่ดังที่กล่าวมาข้างต้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4
การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงผลการทดลองของเหตุการณ์สองเหตุการณ์ นั่นคือ ข้อมูลไม่สมดุล และข้อมูลมีแพตเทิร์นที่เปลี่ยนไป ด้วยแบบจำลองเสริมกำลัง DQN และ Policy gradient เปรียบเทียบกับแบบจำลองการเรียนรู้แบบมีการชี้นำสามแบบจำลอง คือ แรนดอมฟอเรสต์ เอกซ์จีบูสต์ และวิธีเพื่อนบ้านใกล้ที่สุด k ตัว โดยประเมินผลด้วยค่าความแม่นยำ (accuracy) ค่าความเที่ยงตรง (precision) ค่าความครบถ้วน (recall) และค่าเอฟวัน (F1)

4.1 การปรับช่วงข้อมูลของตัวแปร

เนื่องจากข้อมูลที่ใช้ในการวิจัยครั้งนี้ ตัวแปรต้นทั้ง 21 ตัวแปรเป็นข้อมูลเชิงปริมาณทั้งสิ้น และแต่ละตัวแปรมีมาตราส่วนของค่าที่วัดได้แตกต่างกัน รวมถึงค่าของชุดข้อมูลมีลักษณะการกระจายที่ไม่ใช่การแจกแจงแบบปกติ (Normal distribution) นอกจากนี้แบบจำลองที่ใช้ในงานวิจัยนี้มีอัลกอริทึมที่ไวต่อตัวแปรที่มีค่ามากกว่า เช่น วิธีเพื่อนบ้านใกล้ที่สุด k ตัว จะมีการหาระยะห่างของแต่ละตัวแปร ทำให้ตัวแปรที่มีช่วงที่กว้างกว่า เช่น ค่าใช้จ่ายเฉลี่ยต่อเดือน (บาท) มีผลมากกว่าตัวแปรที่มีช่วงแคบกว่า เช่น จำนวนเดือนที่ใช้บริการ ดังนั้นจึงจำเป็นต้องมีการปรับช่วงข้อมูลของตัวแปรต้นให้อยู่ในช่วงเดียวกัน เพื่อให้ตัวแปรมีความสำคัญเท่ากัน โดยใช้ค่าคะแนนมาตรฐาน (z-score) เป็นการปรับการกระจายของข้อมูลให้มีค่าเฉลี่ยเท่ากับศูนย์ และมีค่าเบี่ยงเบนมาตรฐานเท่ากับหนึ่ง โดยการนำค่าเฉลี่ยไปลบ แล้วนำค่าเบี่ยงเบนมาตรฐานไปหารกับค่าที่ต้องการปรับ ดังสมการที่ 20 กำหนดให้ x เป็นค่าของข้อมูลที่ i โดยที่ μ(x) คือค่าเฉลี่ยของชุดข้อมูล x และ **σ**(x) คือค่าเบี่ยงเบนมาตรฐานของชุดข้อมูล x

$$y_i = \frac{x_i - \mu(x)}{\sigma(x)} \tag{20}$$

ดังนั้น ในงานวิจัยนี้ จะทำการปรับค่าของตัวแปรต้นทั้งหมด 21 ตัวแปรด้วยค่าคะแนนมาตรฐานก่อนจะนำไปใช้ในการเรียนรู้แบบจำลองทั้งห้าแบบจำลอง

4.2 เปรียบเทียบสมรรถนะกับแบบจำลองการเรียนรู้ที่มีการชี้นำอื่น ๆ

ในงานวิจัยนี้จะนำแบบจำลองการเรียนรู้ที่มีการชี้นำสามแบบจำลองมาใช้ทำนายการยกเลิกบริการของลูกค้า เพื่อเปรียบเทียบสมรรถนะการทำนายกับแบบจำลองการเรียนรู้เสริมกำลังที่นำเสนอ

ในส่วนนี้จะอธิบายการปรับจูนไฮเปอร์พารามิเตอร์ในแต่ละแบบจำลอง เพื่อหาค่าพารามิเตอร์ที่ดีที่สุด ซึ่งในการปรับจูนนี้ จะใช้ข้อมูลตรวจสอบ (Validation set) จากการแบ่งข้อมูลสอน ด้วยอัตราส่วนข้อมูลตรวจสอบและข้อมูลสอนเท่ากับ 20%:80% โดยจะมีการปรับจูนค่าพารามิเตอร์ของสองแบบจำลอง แบบจำลองที่หนึ่งเป็นแบบจำลองที่ใช้ข้อมูลชุดทีหนึ่งในการเรียนรู้ และแบบจำลองที่สองเป็นแบบจำลองที่ใช้ข้อมูลชุดที่สามที่เกิดจากการรวมข้อมูลชุดที่หนึ่งและข้อมูลชุดที่สองที่มีแพตเทิร์นเปลี่ยนไป

4.2.1 แบบจำลองแรนดอมฟอเรสต์

ในการปรับจูนค่าไฮเปอร์พารามิเตอร์ในแบบจำลองแรนดอมฟอเรสต์ จะทำการหาค่าพารามิเตอร์ที่ดีที่สุดด้วยวิธีค้นหาแบบตะกราม (grid search) เพื่อให้ได้ค่าพารามิเตอร์ทั้ง 5 พารามิเตอร์ ดังตารางที่ 4 ที่ทำให้แบบจำลองมีค่าความแม่นยำมากที่สุด

ตารางที่ 4 การปรับจูนค่าไฮเปอร์พารามิเตอร์ในแบบจำลองแรนดอมฟอเรสต์

พารามิเตอร์	ความหมาย	ค่าพารามิเตอร์ของแบบจำลองที่1	ค่าพารามิเตอร์ของแบบจำลองที่2
n_estimate	จำนวนต้นไม้ทั้งหมดที่จะใช้ในการทำนาย	500	1500
max_dept	ความลึกของการแตกกิ่งของต้นไม้แต่ละต้น	15	10
min_sample_split	จำนวนข้อมูลอย่างน้อยที่สุดที่จะใช้ในการแบ่งโหนดภายใน (internal leaf node)	20	20
min_samples_leaf	จำนวนข้อมูลอย่างน้อยที่สุดที่จะเป็นโหนดใบ (leaf node)	10	50
max_features	จำนวนฟีเจอร์ที่ใช้ในการหาการแบ่งที่ดีที่สุด	5	10

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.2.2 แบบจำลองเอกซ์จีบูสต์

ในการสร้างแบบจำลองเอกซ์จีบูสต์นี้ จะเลือกประเภทบูสต์เตอร์เป็นแบบต้นไม้ และจะทำการปรับจูนค่าไฮเปอร์พารามิเตอร์ด้วยวิธีค้นหาแบบตะกราม จได้ค่าไฮเปอร์พารามิเตอร์ดังตารางที่ 5

ตารางที่ 5 การปรับจูนค่าไฮเปอร์พารามิเตอร์ในแบบจำลองเอกซ์จีบูสต์

พารามิเตอร์	ความหมาย	ค่าพารามิเตอร์ของแบบจำลองที่1	ค่าพารามิเตอร์ของแบบจำลองที่ 2
max_dept	ความลึกการแตกกิ่งของต้นไม้แต่ละต้น	5	3
min_child_weight	ผลรวมของน้ำหนักของข้อมูลทุกตัวที่ต้องการในการแบ่งแต่ละโหนดลูก	6	6
gamma	ค่าของความผิดพลาดที่ลดลงแล้วจะทำการแบ่งโหนดใหม่	0.4	0.0
subsample	สัดส่วนข้อมูลที่จะใช้สร้างต้นไม้ใหม่	0.85	0.85
colsample_bytree	สัดส่วนของฟีเจอร์ที่จะใช้สร้างต้นไม้	0.85	0.90
reg_alpha	น้ำหนักของ แอลวัน เรกูลาไรเซชัน (L1 regularization)	0.00001	0.00001

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.2.3 แบบจำลองวิธีหาเพื่อนบ้านใกล้ที่สุด k ตัว

แบบจำลองวิธีหาเพื่อนบ้านใกล้ที่สุด k ตัว เป็นวิธีที่ทำการวัดระยะห่างแต่ละตัวแปรของข้อมูลที่เราต้องการทำนาย กับชุดข้อมูลสอน โดยอัลกอริทึมที่ใช้ในการหาระยะห่างที่ใกล้ที่สุดนั้นมีหลายอัลกอริทึม เช่น อัลกอริทึมบอล ทรี (Ball Tree algorithm) อัลกอริทึมเคดี ทรี (KD Tree algorithm) และบรูทฟอร์ส (Brute Force) ซึ่งในที่นี้จะให้คอมพิวเตอร์เลือกอัลกอริทึมที่เหมาะสมให้เองโดยพิจารณาจากข้อมูลที่ใช้ในการเรียนรู้ และจะทำการปรับจูนไฮเปอร์พารามิเตอร์ 3 พารามิเตอร์ด้วยวิธีตะกราม เพื่อหาค่าที่ทำให้แบบจำลองมีความแม่นยำมากที่สุด ดังตารางที่ 6

ตารางที่ 6 การปรับจูนค่าไฮเปอร์พารามิเตอร์ในแบบจำลองวิธีหาเพื่อนบ้านใกล้ที่สุด k ตัว

พารามิเตอร์	ความหมาย	ค่าพารามิเตอร์ของแบบจำลองที่ 1	ค่าพารามิเตอร์ของแบบจำลองที่ 2
leaf_size	จำนวนข้อมูลในแต่ละกลุ่มในอัลกอริทึมบอล ทรี และเคดี ทรี	20	30
n_neighbors	จำนวนเพื่อนบ้านที่จะพิจารณา	41	41
p	เลขชี้กำลังของสูตรหาระยะทางของมิงค็อฟสกี (minkowski)	1	1

4.3 ผลการทดลอง

สำหรับผลการทดลองของเหตุการณ์ที่ข้อมูลมีแพตเทิร์นเปลี่ยนไป นั่นคือ ลูกค้ามีการเปลี่ยนสถานะหรือตัวแปรป้าย จากยกเลิกบริการเป็นไม่ยกเลิกบริการ และจากยกเลิกบริการเป็นไม่ยกเลิกบริการ ในการวัดความสามารถของการทำนาย จะทำการวัดผลบนข้อมูลสองชุด ประกอบด้วย ข้อมูลชุดที่หนึ่งซึ่งเป็นข้อมูลปกติ และข้อมูลชุดที่สองที่มีแพตเทิร์นข้อมูลเปลี่ยนไป โดยอัตราส่วนของลูกค้ายกเลิกและไม่ยกเลิกบริการของทั้งสองชุดข้อมูลค่อนข้างสมดุล ดังนั้น สามารถเปรียบเทียบค่าความแม่นยำของแต่ละแบบจำลองได้ จากตารางที่ 7 แสดงการเปรียบเทียบสมรรถนะระหว่างแบบจำลองการเรียนรู้ที่มีการชี้นำ และแบบจำลองการเรียนรู้เสริมกำลังในข้อมูลทั้งสองชุด โดยเมื่อพิจารณาสมรรถนะของแบบจำลองในข้อมูลชุดที่หนึ่ง จะเห็นว่าแบบจำลองการเรียนรู้เสริมกำลัง Policy gradient ให้ค่าความแม่นยำมากที่สุด โดยมีค่าเท่ากับ 86.3% รองลงมาคือแบบจำลอง DQN ที่ให้ค่าความแม่นยำเท่ากับ 83.6% เปรียบเทียบกับแบบจำลองการเรียนรู้ที่มีการชี้นำทั้งสามแบบจำลองให้ค่าความแม่นยำที่ใกล้เคียงกัน โดยแบบจำลองแรนดอมฟอเรสต์มีความแม่นยำมากที่สุดเท่ากับ 61.4% รองลงมาคือ เอกซ์จีบูสต์ 61.2% และวิธีเพื่อนบ้านใกล้ที่สุด k ตัว 59.3% ตามลำดับ นอกจากนี้ ในงานการทำนายลูกค้ายกเลิกบริการ ค่าความครบถ้วนก็เป็นอีกค่าที่สำคัญ เมื่อพิจารณาค่าความครบถ้วน พบว่า DQN ให้ค่าความครบถ้วนสูงที่สุดเท่ากับ 90.3% รองลงมาคือ Policy gradient เท่ากับ 84.7% และเมื่อพิจารณาในข้อมูลชุดที่มีแพตเทิร์นเปลี่ยนไป เห็นได้ว่าทุกแบบจำลองมีค่าทั้งสี่ลดลงเมื่อเทียบกับสมรรถนะการทำนายของข้อมูลชุดที่หนึ่ง แต่แบบจำลองการเรียนรู้เสริมกำลัง Policy gradient ยังคงให้ค่าความแม่นยำสูงที่สุด เท่ากับ 83.7% รองลงมาคือ DQN ที่มีค่าความแม่นยำเท่ากับ 73.3% ในส่วนของแบบจำลองการเรียนรู้ที่มีการชี้นำทั้ง 3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

แบบจำลองมีค่าความแม่นยำลดลงในข้อมูลชุดที่มีแพตเทิร์นเปลี่ยนไป โดยแบบจำลองเอกซ์จีบูสต์ให้ค่ามากที่สุดเท่ากับ 55.6% รองลงมาคือแรนดอมฟอเรสต์ 55.5% และ วิธีหาเพื่อนบ้านใกล้ที่สุด k ตัว 53.3% ตามลำดับ

ตารางที่ 7 สรุปเปรียบเทียบสมรรถนะระหว่างแบบจำลองด้วยค่าความแม่นยำ ความเที่ยงตรง ความครบถ้วน และเอฟวันของข้อมูลชุดที่หนึ่ง และข้อมูลชุดที่มีแพตเทิร์นเปลี่ยนไป

แบบจำลอง	ข้อมูลชุดทีหนึ่ง				ข้อมูลชุดที่มีแพตเทิร์นเปลี่ยนไป			
	ความแม่นยำ	ความเที่ยงตรง	ความครบถ้วน	เอฟวัน	ความแม่นยำ	ความเที่ยงตรง	ความครบถ้วน	เอฟวัน
Random forest	61.4%	59.6%	63.0%	61.3%	55.5%	55.4%	48.8%	51.9%
XGBoost	61.2%	59.8%	61.3%	60.5%	55.6%	55.4%	50.3%	52.8%
kNN	59.3%	58.8%	54.1%	56.4%	53.3%	52.9%	46.1%	49.3%
DQN	83.6%	78.9%	**90.3%**	84.2%	73.3%	66.2%	**95.2%**	78.1%
Policy gradient	**86.3%**	**86.7%**	84.7%	**85.7%**	**83.7%**	**83.5%**	84.0%	**83.8%**

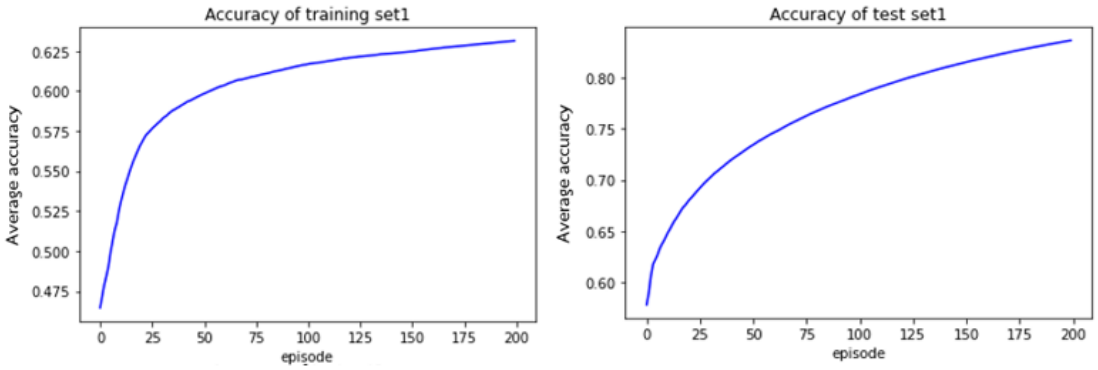
*****หมายเหตุ ตัวหนาคือค่าที่มากที่สุดในคอลัมน์

นอกจากนี้ แบบจำลอง DQN และ Policy gradient ยังเป็นแบบจำลองที่มีความสามารถในการปรับตัวกับข้อมูลได้ดี ดังแสดงในภาพที่ 15 - 18 จากภาพที่ 15 พบว่า ค่าความแม่นยำเฉลี่ยในแต่ละรอบการเรียนรู้ของแบบจำลอง DQN ข้อมูลสอนและข้อมูลทดสอบของข้อมูลชุดทีหนึ่ง พิจารณาภาพที่ 15 (ซ้าย) กราฟมีแนวโน้มเพิ่มขึ้นเรื่อย ๆ โดยจะมีการเรียนรู้และปรับตัวอย่างรวดเร็วในช่วงแรก ในรอบที่ 1 ระบบ (agent) จะยังทำนายได้ไม่ค่อยแม่นยำ ทำให้ค่าความแม่นยำมีค่าประมาณ 45% แต่เมื่อระบบได้ทำการเรียนรู้และปรับค่าน้ำหนักของโครงข่ายจากค่าผลรางวัลที่ได้ในแต่ละรอบ จะเห็นว่าในรอบที่ 25 ค่าความแม่นยำมีค่าเพิ่มขึ้นเป็น 57.5% และยังคงเพิ่มขึ้นเรื่อย ๆ จนถึงรอบที่ 200 ค่าความแม่นยำเท่ากับ 62.5% เช่นเดียวกับค่าความแม่นยำของข้อมูลทดสอบของข้อมูลชุดทีหนึ่ง ดังภาพที่ 16 (ขวา) ในรอบที่ 1 ของการเรียนรู้ แบบจำลอง DQN สามารถทำนายได้แม่นยำประมาณ 57% เมื่อเรียนรู้จนถึงรอบที่ 25 แบบจำลองสามารถทำนายได้แม่นยำสูงถึง 70% และปรับตัวสูงขึ้นเรื่อย ๆ จนรอบที่ 200 สามารถทำนายได้แม่นยำสูงถึง 83.6% เมื่อให้แบบจำลองเรียนรู้ข้อมูลที่มีแพตเทิร์นเปลี่ยนไป สำหรับภาพที่ 16 (ซ้าย) แสดงสมรรถนะการทำนายข้อมูลสอน จะเห็น

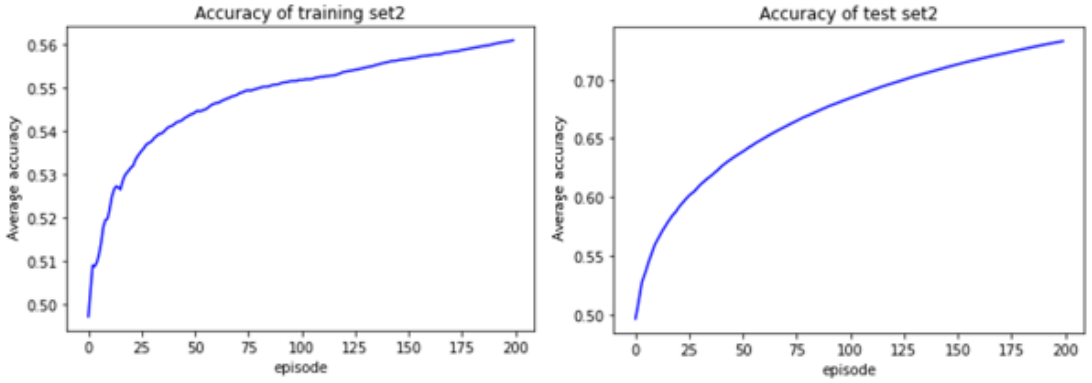
จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ว่าในรอบที่ 1 สมรรถนะการทำนายของแบบจำลอง DQN ลดลงอยู่ที่ประมาณ 50% และในรอบที่ 25 สามารถทำนายได้แม่นยำเพิ่มขึ้นจนถึง 53.5% และปรับตัวเพิ่มขึ้นเรื่อย ๆ จนรอบที่ 200 ทำนายได้แม่นยำถึง 56% เช่นเดียวกับค่าความแม่นยำของข้อมูลทดสอบของข้อมูลชุดที่มีแพตเทิร์นเปลี่ยนไป ดังภาพที่ 16 (ขวา)

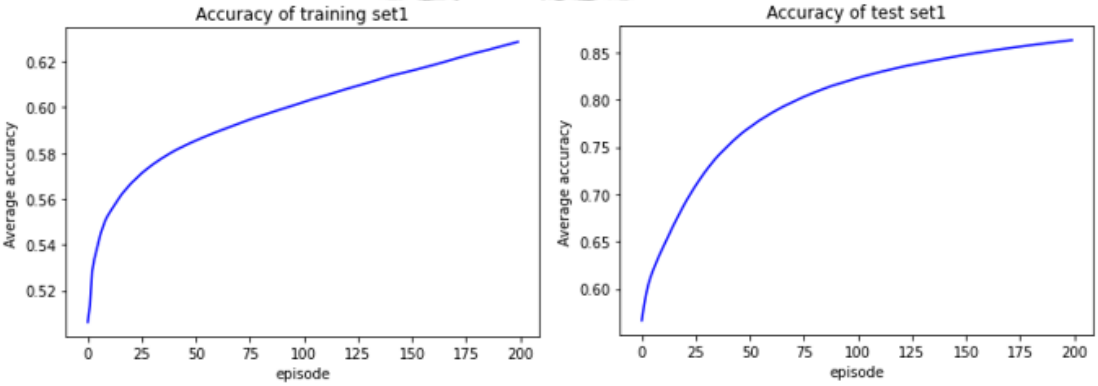
พิจารณาภาพที่ 17 และ ภาพที่18 แสดงค่าความแม่นยำเฉลี่ยแต่ละรอบการเรียนรู้ของแบบจำลอง Policy gradient ซึ่งแสดงให้เห็นถึงการปรับตัวได้ดีกับข้อมูลที่เปลี่ยนแปลงไป จากภาพที่ 17 (ซ้าย) ในรอบที่ 1 ของการเรียนรู้ แบบจำลองสามารถทำนายข้อมูลชุดที่หนึ่งได้แม่นยำเท่ากับ 50% เมื่อเรียนรู้จนถึงรอบที่ 25 แบบจำลองสามารถทำนายได้แม่นยำเพิ่มขึ้นเป็น 58% และเพิ่มขึ้นอย่างต่อเนื่องจนรอบที่ 200 ที่ให้ค่าความแม่นยำเท่ากับ 63% และในภาพที่ 17 (ขวา) ได้ทำการทดสอบความแม่นยำของแบบจำลอง Policy gradient กับข้อมูลทดสอบของชุดหนึ่ง กราฟจะมีลักษณะเช่นเดียวกัน คือในช่วงแรกของการเรียนรู้ในรอบที่ 1 มีค่าความแม่นยำเท่ากับ 56% และปรับตัวเพิ่มมากขึ้น จนรอบที่ 200 สามารถทำนายได้แม่นยำ 86.3% เมื่อพิจารณาสมรรถนะของแบบจำลองในชุดข้อมูลที่มีแพตเทิร์นเปลี่ยนไป ภาพที่ 18 แสดงถึงแบบจำลองสามารถปรับตัวและทำนายได้ดี จากภาพที่ 18 (ซ้าย) กราฟมีลักษณะเกือบเป็นเส้นตรง โดยเริ่มต้นสามารถทำนายได้แม่นยำเท่ากับ 50% และทำนายได้แม่นยำขึ้นเรื่อย ๆ จนรอบที่ 200 สามารถทำนายได้แม่นยำ 62% และเมื่อวัดสมรรถนะกับข้อมูลทดสอบดังภาพที่ 18 (ขวา) แบบจำลองยังคงสามารถทำนายได้ดี โดยเริ่มต้นที่ความแม่นยำเท่ากับ 50% และเพิ่มขึ้นจนเท่ากับ 83.7% ในรอบที่ 200



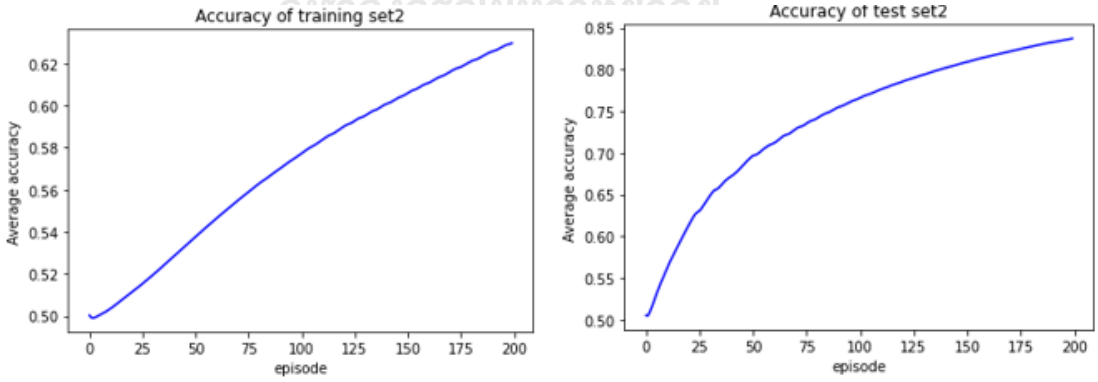
ภาพที่ 15 (ซ้าย) ค่าความแม่นยำเฉลี่ยของแบบจำลอง DQN ในแต่ละรอบการเรียนรู้ของข้อมูลสอนชุดที่หนึ่ง (ขวา) ค่าความแม่นยำเฉลี่ยของข้อมูลทดสอบชุดที่หนึ่ง



ภาพที่ 16 (ซ้าย) ค่าความแม่นยำเฉลี่ยของแบบจำลอง DQN ในแต่ละรอบการเรียนรู้ของข้อมูลสอนชุดที่สอง (ขวา) ค่าความแม่นยำเฉลี่ยของข้อมูลทดสอบชุดที่สอง



ภาพที่ 17 (ซ้าย) ค่าความแม่นยำเฉลี่ยของแบบจำลอง Policy gradient ในแต่ละรอบการเรียนรู้ของข้อมูลสอนชุดที่หนึ่ง (ขวา) ค่าความแม่นยำเฉลี่ยของข้อมูลทดสอบชุดที่หนึ่ง



ภาพที่ 18 (ซ้าย) ค่าความแม่นยำเฉลี่ยของแบบจำลอง Policy gradient ในแต่ละรอบการเรียนรู้ของข้อมูลสอนชุดที่สอง (ขวา) ค่าความแม่นยำเฉลี่ยของข้อมูลทดสอบชุดที่สอง

บทที่ 5
สรุปผลการวิจัย

5.1 สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้ได้นำเสนอแบบจำลองการเรียนรู้เสริมกำลังมาใช้ในการทำนายการยกเลิกบริการของลูกค้า ที่มีลักษณะเป็นโจทย์การเรียนรู้แบบมีผู้สอน โดยได้ทำการทดลองกับข้อมูลที่มีแพตเทิร์นของข้อมูลเปลี่ยนไปจากเดิม โดยข้อมูลที่นำมาใช้ในงานวิจัยนี้ เป็นข้อมูลการใช้งานของลูกค้าโทรคมนาคมแห่งหนึ่ง ซึ่งมีอัตราส่วนของลูกค้ายกเลิกบริการ และไม่ยกเลิกบริการเท่ากับ 48%:52% ซึ่งเป็นข้อมูลที่สมดุล ผู้วิจัยได้ทำการจำลองข้อมูลขึ้นมาหนึ่งชุดที่มีแพตเทิร์นของข้อมูลเปลี่ยนไปจากเดิม นั่นคือลูกค้ามีสถานะที่เปลี่ยนไป เช่น ลูกค้ามีสถานะเป็นไม่ยกเลิกบริการ แต่เมื่อเวลาผ่านไปลูกค้าจะมีสถานะเปลี่ยนเป็นยกเลิกบริการ โดยในงานวิจัยนี้ได้ใช้แบบจำลองการเรียนรู้เสริมกำลัง 2 แบบจำลอง คือ DQN และ Policy gradient เปรียบเทียบกับแบบจำลองการเรียนรู้แบบมีการชี้นำ 3 แบบจำลอง คือ แรนดอมฟอเรสต์ เอกซ์จีบูสต์ และ วิธีหาเพื่อนบ้านใกล้ที่สุด k ตัว ผลการทดลองพบว่า แบบจำลอง Policy gradient ให้ค่าความแม่นยำสูงที่สุดทั้งข้อมูลชุดทีหนึ่ง และข้อมูลชุดสองที่มีข้อมูลที่แพตเทิร์นเปลี่ยนไป โดยมีความแม่นยำเท่ากับ 86.3% และ 83.7% ตามลำดับ รองลงมาคือ DQN ให้ค่าความแม่นยำเท่ากับ 83.6% และ 73.3% ตามลำดับ สำหรับแบบจำลองการเรียนรู้ของเครื่องอีก 3 แบบจำลอง แบบจำลองแรนดอมฟอเรสต์ และเอกซ์จีบูสต์ให้ค่าความแม่นยำที่ใกล้เคียงกัน นั่นคือ 61.4% และ 61.2% สำหรับข้อมูลชุดที่หนึ่ง และ 55.5% และ 55.6% ในข้อมูลชุดที่สอง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากผลการทดลอง พบว่าแบบจำลองการเรียนรู้แบบเสริมกำลัง สามารถเรียนรู้ข้อมูลที่แพตเทิร์นเปลี่ยนไปจากเดิมได้ดีกว่าแบบจำลองการเรียนรู้แบบมีการชี้นำ เนื่องจากแบบจำลองการเรียนรู้แบบมีการชี้นำเป็นการเรียนรู้ที่ไม่โต้ตอบ (passive learning) นั่นคือ การที่แบบจำลองถูกสร้างกับข้อมูลชุดหนึ่งแล้ว เมื่อข้อมูลมีลักษณะที่เปลี่ยนไปจากเดิม แบบจำลองเดิมจะไม่สามารถทำนายข้อมูลใหม่นั้นได้อย่างแม่นยำ แต่แบบจำลองการเรียนรู้เสริมกำลังเป็นการเรียนรู้แบบเชิงรุก (active learning) โดยจะมีระบบ (agent) ที่คอยสื่อสารกับสิ่งแวดล้อมอยู่ตลอดเวลา และระบบจะมีการเรียนรู้ซ้ำ ๆ โดยแต่ละรอบในการเรียนรู้ ระบบจะได้ค่าผลรางวัลว่าการกระทำที่ตัดสินใจทำไปนั้นเป็นการกระทำที่ดีหรือไม่ แล้วนำค่าผลรางวัลนั้นไปใช้ในการปรับแบบจำลองให้มีความแม่นยำการทำนายที่ดียิ่งขึ้น ดังนั้น ในแต่ละรอบการเรียนรู้ระบบจะสามารถทำนายได้แม่นยำมากยิ่งขึ้น และเมื่อมีข้อมูลที่มี

แพตเทิร์นเปลี่ยนไปเข้ามาเป็นข้อมูลใหม่ ระบบจะทำการเรียนรู้ข้อมูลนั้นซ้ำ ๆ และมีการนำค่าผลรางวัลที่ได้มาปรับให้สามารถทำนายได้แม่นยำมากยิ่งขึ้นนั่นเอง

5.2 แนวทางงานวิจัยในอนาคต

เนื่องจากงานวิจัยนี้ได้ทำการทดลองกับข้อมูลที่ไม่มีความสัมพันธ์กัน นั่นคือข้อมูลลูกค้าหนึ่งคน จะถูกเก็บเป็นข้อมูลหนึ่งแถว ไม่เป็นแถวลำดับต่อเนื่องกัน แต่การเรียนรู้เสริมกำลังยังสามารถจัดการกับข้อมูลที่มีลักษณะเป็นลำดับต่อเนื่องกันได้ดี ดังนั้น งานวิจัยในอนาคตอาจทำการทดลองกับข้อมูลอนุกรมเวลา หรือข้อมูลที่มีการกระทำเป็นลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บรรณานุกรม

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

[1] Wertz, J., Don’t spend 5 times more attracting new customers, nurture the existing ones. [online]. 2561. URL: https://www.forbes.com/sites/jiawertz/2018/09/12/dont-spend-5-times-more-attracting-new-customers-nurture-the-existing-ones/#519f48e75a8e

[2] Lopez-Martin, M., Carro, B. and Sanchez-Esguevillas, A. Application of deep reinforcement learning to intrusion detection for supervised problems. In: Expert Systems with Applications, volume141, 2020.

[3] Sutton, R.S., and Barto, A.G., Reinforcement Learning: An Introduction. second edition. Cambridge, MA: The MIT Press, 2018.

[4] Rai, S., Khandelwal, N. and Boghey, R., Analysis of customer churn prediction in telecom sector using cart algorithm. In: Advances in Intelligent Systems and Computing, volume1045, p.457-466. 2019.

[5] Keramati, A., Jafari-Marandi, R., Aliannejad, M., Ahmadian, I. and Mozaffari, M. Improved churn prediction in telecommunication industry using data mining techniques. In: Applied Soft Computing, volume24, p.994-1012, 2014.

[6] Cao, S., Liu, W., Chen, Y. and Zhu, X. Deep Learning Based Customer Churn Analysis. In: 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP) Xi'an, China, p.1-6, 2019.

[7] Ahmad, A., Jafar, A., and Aljoumaa, K. Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data., 2019.

[8] Wong, M.L., Seng, K. and Wong, P.K., Cost-sensitive ensemble of stacked denoising autoencoders for class imbalance problems in business domain. In: Expert Systems with Applications, volume141, 2020.

[9] Devrishi., New Usability Rating on Datasets. [online] 2562. URL: https://www.kaggle.com/product-feedback/93922

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียน

ชื่อ-สกุล	เมธาวี ปัญจสุชาติ
วัน เดือน ปี เกิด	26 ตุลาคม 2536
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	ปริญญาตรีจากคณะพาณิชยศาสตร์และการบัญชี สาขาวิชาสถิติ จุฬาลงกรณ์มหาวิทยาลัย
ที่อยู่ปัจจุบัน	96 ซ.เอกชัย42 ถ.เอกชัย แขวงบางคลองพราน เขตบางบอน กรุงเทพมหานคร 10150
ผลงานตีพิมพ์	Applying Reinforcement Learning for Customer Churn Prediction, 2020, ICCEE, Jul 2020

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University