ฟังก์ชันสูญเสียสำหรับการเรียนรู้ข้อมูลที่ไม่สมดุลของแบบจำลองโครงข่ายประสาทเทียมเชิงลึก
Hybrid Loss for Learning Imbalanced Data

ธนวัฒน์ หลอดแก้ว
Thanawat Lodkaew
59070071

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ภาคเรียนที่ 1 ปีการศึกษา 2562

ฟังก์ชันสูญเสียสำหรับการเรียนรู้ข้อมูลที่ไม่สมดุลของแบบจำลองโครงข่ายประสาทเทียมเชิงลึก
Hybrid Loss for Learning Imbalanced Data

โดย

ธนวัฒน์ หลอดแก้ว
รหัสประจำตัว 59070071

อาจารย์ที่ปรึกษา
รองศาสตราจารย์ ดร. กิติ์สุชาต พสุภา

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ภาคเรียนที่ 1 ปีการศึกษา 2562

Hybrid Loss for Learning Imbalanced Data

Thanawat Lodkaew

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
BACHELOR OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
1/2019

ใบรับรองปริญญานิพนธ์ ประจำปีการศึกษา 2562
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง **ฟังก์ชันสูญเสียสำหรับการเรียนรู้ข้อมูลที่ไม่สมดุลของแบบจำลองโครงข่ายประสาทเทียมเชิงลึก**
HYBRID LOSS FOR LEARNING IMBALANCED DATA

ผู้จัดทำ

1. ธนวัฒน์ หลอดแก้ว รหัสประจำตัว 59070071

รองศาสตราจารย์ ดร. กิติ์สุชาต พสุภา อาจารย์ที่ปรึกษา

(...)

ใบรับรองโครงงาน (PROJECT)

เรื่อง

ฟังก์ชันสูญเสียสำหรับการเรียนรู้ข้อมูลที่ไม่สมดุลของแบบจำลอง โครงข่ายประสาทเทียมเชิงลึก

Hybrid Loss for Learning Imbalanced Data

ธนวัฒน์ หลอดแก้ว รหัสประจำตัว 59070071

ขอรับรองว่ารายงานฉบับนี้ ข้าพเจ้าไม่ได้คัดลอกมาจากที่ใด
รายงานฉบับนี้ได้รับการตรวจสอบและอนุมัติให้เป็นส่วนหนึ่งของ
การศึกษาวิชาโครงงาน หลักสูตรวิทยาศาสตรบัณฑิต (เทคโนโลยีสารสนเทศ)
ภาคเรียนที่ 1 ปีการศึกษา 2562

....
ธนวัฒน์ หลอดแก้ว

ชื่อรายงาน ฟังก์ชันสูญเสียสำหรับการเรียนรู้ข้อมูลที่ไม่สมดุลของแบบจำลองโครงข่ายประสาทเทียมเชิงลึก
ชื่อนักศึกษา ธนวัฒน์ หลอดแก้ว
รหัสนักศึกษา 59070071
สาขาวิชา เทคโนโลยีสารสนเทศ
อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร. กิติ์สุชาต พสุภา
ปีการศึกษา 2562

บทคัดย่อ

ปัญหาความไม่สมดุลของข้อมูลเป็นเรื่องที่ถูกหยิบขึ้นมาศึกษาอย่างแพร่หลาย แบบจำลองโครงข่ายประสาทเทียมเชิงลึกสามารถให้ผลการทำงานที่ย่ำแย่เมื่อมันเรียนรู้จากชุดข้อมูลที่ไม่สมดุลกัน งานวิจัยนี้ได้นำเสนอฟังก์ชันสูญเสียแบบผสมผสานที่จะช่วยให้แบบจำลองสามารถเรียนรู้ข้อมูลที่ไม่สมดุลอย่างมีประสิทธิภาพ ฟังก์ชันสูญเสียนี้เป็นการผสมผสานกันระหว่างแนวคิดการคำนวณค่าสูญเสียของสองฟังก์ชันสูญเสีย ที่ซึ่งทั้งสองฟังก์ชันเป็นฟังก์ชันสูญเสียที่ถูกออกแบบมาเพื่อจัดการกับปัญหาความไม่สมดุลของข้อมูลอยู่แล้ว และมีแนวคิดการแก้ปัญหาที่ต่างกันและน่าสนใจ ผู้วิจัยจึงได้เสนอที่จะรวมแนวคิดของทั้งสองฟังก์ชันดังกล่าวเข้าด้วยกัน เพื่อที่จะทำให้ประสิทธิภาพของการเรียนรู้ของแบบจำลองนั้นดีขึ้น ฟังก์สูญเสียแบบผสมผสานที่นำเสนอถูกทดสอบกับชุดข้อมูลที่หลากหลาย และผลการทดลองแสดงให้เห็นว่าฟังก์ชันสูญเสียที่นำเสนอสามารถเพิ่มประสิทธิภาพของแบบจำลองได้เหนือกว่าฟังก์ชันสูญเสียอื่น ๆ

Project Title	Hybrid Loss for Learning Imbalanced Data
Name	Thanawat Lodkaew
Student ID	59070071
Department	Information Technology
Advisor	Assoc. Prof. Dr. Kitsucart Pasupa
Year	2019

Abstract

Classification of imbalanced data is extremely common in practice, and this problem has been widely studied in classical machine learning. A deep neural network (DNN) model produced from an imbalanced data set is likely to be biased towards the majority class and show inferior classification accuracy on the minority class. This work aims at inventing a new loss for learning imbalanced data. This loss is hybridized by two well-performed loss functions, mean false error (MFE) and focal loss (FL). The two loss functions are designed to combat the imbalance problem, and each of them has its own advantage. Hence, we propose to hybridize the two losses and redefine as a hybrid loss that applies the calculation procedures of MFE's total loss to the focal loss. The proposed loss function is tested with several imbalanced datasets, and our experimental results show that it can overcome the existing loss functions.

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยยความกรุณาจาก อาจารย์ ดร. กิติ์สุชาต พสุภา อาจารย์ที่ปรึกษาปริญญานิพนธ์ ที่ได้ให้คาปรึกษาแนะนำชี้แนะแนวทางในการศึกษาค้นคว้า ตลอดจนช่วยแก้ไขข้อบกพร่องต่าง ๆ ในการทาปริญญานิพนธ์มาโดยตลอดจนโครงงานนี้สำเร็จ ลุล่วงด้วยดี คณะผู้จัดทำจึงขอกราบขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณคณาจารย์ทุกท่านที่ช่วยอบรมวิชาความรู้และให้แนวคิดต่าง ๆ ที่เป็นประโยชน์ และคณะเทคโนโลยีสารสนเทศสถาบัยเทคโนโลยีเจ้าคุณทหารลาดกระบัง ที่เอื้อเฟื้อสถานที่ วัสดุอุปกรณ์ต่าง ๆ สาหรับจัดทำโครงงาน และขอขอบคุณสื่อการสอนออนไลน์ บทความหรือวิจัยต่าง ๆ ที่ให้ข้อมูลที่เป็นประโยชน์ต่อการจัดทำโครงงานครั้งนี้เป็นอย่างมาก

ขอขอบคุณ เพื่อน รุ่นพี่ภายในคณะเทคโนโลยีสารสนเทศ และผู้มีส่วนร่วมเกี่ยวข้องที่ได้ให้ คำปรึกษาและให้ความช่วยเหลือที่ดีมาตลอด สุดท้ายนี้ขอขอบพระคุณบิดามารดาและครอบครัวที่คอยให้ คำปรึกษาในเรื่องต่าง ๆ รวมทั้งเป็นกำลังใจช่วยเป็นแรงผลักดันให้ปริญญานิพนธ์นี้สำเร็จลุล่วงด้วยดี

คณะผู้จัดทำจึงขอกราบขอบพระคุณเป็นอย่างยิ่ง ไว้ ณ โอกาสนี้

ธนวัฒน์ หลอดแก้ว
ผู้จัดทำรายงาน
วันที่ 10 พฤศจิกายน พ.ศ. 2561

สารบัญ

สารบัญตาราง

หน้า

สารบัญภาพ

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

ในด้านการเรียนรู้ของเครื่องจักร (Machine Learning) ความไม่สมดุลกันของข้อมูล หมายถึง การที่จำนวนตัวอย่างของแต่ละกลุ่มข้อมูลมีจำนวนไม่เท่ากัน ซึ่งความไม่สมดุลกันของข้อมูลนี้ถูกนิยามให้เป็นปัญหาในการจัดกลุ่มข้อมูล (Classification) สาเหตุที่ความไม่สมดุลกันของข้อมูลเป็นปัญหา คือ อัลกอริทึมการจัดกลุ่ม จะทำงานได้อย่างมีประสิทธิภาพก็ต่อเมื่อจำนวนตัวอย่างของแต่ละกลุ่มข้อมูลมีจำนวนที่เท่าหรือใกล้เคียงกัน เมื่อมีความไม่สมดุลกันของข้อมูลจะทำให้การทำงานของอัลกอริทึมมีประสิทธิภาพด้อยลง ซึ่งอาจจะด้อยลงจนไม่สามารถจัดกลุ่มข้อมูลได้เลย

การจัดกลุ่มข้อมูลที่มีจำนวนตัวอย่างของแต่ละกลุ่มไม่สมดุลกันเป็นเรื่องธรรมดาอย่างมากในทางปฏิบัติ เนื่องจากข้อมูลที่เกิดขึ้นล้วนแต่ไม่สามารถคาดเดาได้อย่างแน่นอนว่าจำนวนตัวอย่างของแต่ละกลุ่มจะสมดุลกัน อีกทั้งข้อมูลส่วนใหญ่ยังมีลักษณะที่มีจำนวนตัวอย่างของแต่ละกลุ่มไม่สมดุลกัน เช่น ในระหว่างวัวอยู่ในช่วงเป็นสัด ช่วงเวลาที่วัวแสดงพฤติกรรมเป็นสัดจะมีจำนวนน้อยกว่าช่วงเวลาที่วัวไม่แสดงพฤติกรรมเป็นสัด เป็นต้น ในด้านการเรียนรู้ของเครื่องจักร มีความเป็นไปได้ว่าตัวจัดกลุ่มข้อมูลที่ถูกสร้างขึ้นจากชุดข้อมูลที่มีจำนวนตัวอย่างของแต่ละกลุ่มไม่สมดุลกันจะมีความลำเอียงในการจัดกลุ่ม กล่าวคือ มีโอกาสสูงที่ตัวจัดกลุ่มจะระบุว่าข้อมูลเป็นกลุ่มส่วนมาก (Majority Class) มากกว่าเป็นกลุ่มส่วนน้อย (Minority Class) ซึ่งเป็นผลทำให้การระบุข้อมูลเป็นกลุ่มส่วนน้อยมีความแม่นยำที่ต่ำกว่ามาตรฐาน ซึ่งความแม่นยำในการระบุข้อมูลเป็นแต่ละกลุ่มควรจะเท่าหรือใกล้เคียงกัน

ที่ผ่านมาได้มีการศึกษาเกี่ยวกับปัญหาการจัดกลุ่มข้อมูลในลักษณะนี้อย่างกว้างขวาง และได้แสดงให้เห็นถึงความสำคัญของปัญหาของการจัดกลุ่มข้อมูลที่มีจำนวนตัวอย่างของแต่ละกลุ่มไม่สมดุลกัน ซึ่งทำให้ประสิทธิภาพการจัดกลุ่มข้อมูลมีความแม่นยำที่ต่ำ ดังนั้นปัญหานี้จำเป็นต้องถูกจัดการ[1] เพื่อที่จะแก้ปัญหาการจัดกลุ่มข้อมูลที่มีจำนวนตัวอย่างของแต่ละกลุ่มไม่สมดุลกัน ได้มีเทคนิคเกิดขึ้นมากมาย โดยสามารถแบ่งเทคนิคการแก้ปัญหาได้ 2 ระดับ คือ (1) ระดับข้อมูล (Data-Level) ที่ซึ่งเป็นการแก้ปัญหาโดยการจัดการข้อมูลก่อนที่จะถูกนำไปประมวลในกระบวนการจัดกลุ่มข้อมูล โดยการสุ่มเพิ่มจำนวนตัวอย่างข้อมูล (Over-Sampling) และการสุ่มลดจำนวนตัวอย่างข้อมูล (Under-Sampling) เป็นเทคนิคในการแก้ปัญหาในระดับข้อมูล เทคนิคการแก้ปัญหาในระดับนี้เป็นการแก้ปัญหาแบบเบื้องต้นที่สามารถดำเนินการได้ง่าย อย่างไรก็ตามการสุ่มเพิ่มจำนวนตัวอย่างข้อมูลสามารถทำให้เกิดปัญหา Overfitting ตามมาได้อย่างง่ายดาย ในทางเดียวกันการสุ่มลดจำนวนตัวอย่างข้อมูลอาจจะเป็นการกำจัดสารสนเทศที่เป็นประโยชน์ต่อการจัดกลุ่มข้อมูลออกไป (2) ระดับตัวจัดกลุ่ม (Classifier-Level) ที่ซึ่งเป็นการแก้ปัญหาโดยการจัดการอัลกอริทึมการจัดกลุ่ม โดยการ

ทำเทรสโช (Thresholding) การเรียนรู้แบบความเสียหายที่รู้สึกได้ง่าย (Cost-Sensitive Learning) การจัดกลุ่มข้อมูลแบบหนึ่งกลุ่ม (One-Class Classification) และการผนวกกันของหลายเทคนิค อย่างไรก็ตามเทคนิคเหล่านี้มียังไม่สามารถแก้ปัญหาได้อย่างมีประสิทธิภาพในทุก ๆ ชุดข้อมูล กล่าวคือ เทคนิคสามารถให้ความแม่นยำในการจัดกลุ่มข้อมูลแต่ละกลุ่มได้อย่างน่าพอใจสำหรับชุดข้อมูล A แต่ไม่สามารถทำได้อย่างมีประสิทธิภาพสำหรับชุดข้อมูล B เป็นต้น ดังนั้นเทคนิคใหม่ที่จะสามารถการแก้ปัญหาการจัดกลุ่มข้อมูลที่ไม่สมดุลกันได้อย่างมีประสิทธิภาพ และปรับเข้าได้กับทุกชุดข้อมูลจำเป็นต้องถูกคิดค้นขึ้น

ในปัจจุบันโครงข่ายประสาทเทียมเชิงลึก (Deep Neural Network: DNN) ถูกนำมาใช้ในการจัดกลุ่มข้อมูลอย่างกว้างขวาง จากงานวิจัยที่ผ่านมาได้แสดงให้เห็นว่าแบบจำลอง DNN สามารถจัดกลุ่มได้อย่างมีประสิทธิภาพ อย่างไรก็ตามเมื่อข้อมูลมีความไม่สมดุลกันแบบจำลอง DNN จะไม่สามารถจัดกลุ่มได้อย่างที่ควรจะเป็น[1;2] เนื่องจากการเรียนรู้ที่ไม่มีประสิทธิภาพ สำหรับแบบจำลอง DNN เมื่อเรียนรู้จากข้อมูลที่ไม่สมดุลกัน อาจจะทำให้เกิดความลำเอียงในการเรียนรู้ขึ้น ที่ซึ่งแบบจำลองจะมุ่งเรียนรู้ข้อมูลจากกลุ่มข้อมูลส่วนมากมากกว่าเรียนรู้ข้อมูลจากกลุ่มข้อมูลส่วนน้อย ทำให้แบบจำลองไม่สามารถแยกแยะข้อมูลว่าเป็นของกลุ่มข้อมูลส่วนมากหรือกลุ่มข้อมูลส่วนน้อยได้อย่างมีประสิทธิภาพ

เพื่อที่จะทำให้แบบจำลอง DNN เรียนรู้ข้อมูลที่ไม่สมดุลกันได้อย่างมีประสิทธิภาพ งานวิจัยจึงได้นำเสนอฟังก์ชันสูญเสีย (Loss Function) แบบใหม่ ที่จะทำให้การเรียนรู้ของแบบจำลองมีความสมดุลกัน โดยฟังก์ชันสูญเสียแบบใหม่นี้ถูกดัดแปลงมาจากฟังก์ชันสูญเสียที่มีอยู่แล้ว ที่ซึ่งจะคำนึงถึงข้อได้เปรียบของแต่ละฟังก์ชันสูญเสีย แล้วนำแนวคิดของฟังก์ชันสูญเสียเหล่านั้นมาผสมผสานกัน เพื่อสร้างเป็นฟังก์ชันสูญเสียใหม่

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาการเรียนรู้ของแบบจำลอง DNN ด้วยฟังก์ชันสูญเสียแบบต่าง ๆ
2. เพื่อคิดค้นฟังก์ชันสูญเสียแบบใหม่ที่ทำให้แบบจำลอง DNN สามารถเรียนรู้จากข้อมูลที่ไม่สมดุลได้อย่างมีประสิทธิภาพ

1.3 ขอบเขตการพัฒนาโครงงาน

1. คิดค้นฟังก์ชันสูญเสียแบบใหม่ และทำการทดลองเพื่อเปรียบเทียบประสิทธิภาพของฟังก์ชันสูญเสียแบบใหม่กับฟังก์ชันสูญเสียที่มีอยู่

1.4 ขั้นตอนการดำเนินงาน

1. ศึกษาเกี่ยวกับนิยามของความไม่สมดุลกันของข้อมูลในด้านการจัดกลุ่มข้อมูล
2. ศึกษากระบวนการการเรียนรู้ของแบบจำลอง DNN และการทำงานของฟังก์ชันสูญเสียแบบต่าง ๆ

3. ศึกษางานวิจัยที่เกี่ยวข้อง
4. ตั้งข้อสมมติฐาน
5. ออกแบบการทดลอง
6. เลือกชุดข้อมูล และ Metrics ที่จะใช้ในการทดลอง
7. ดำเนินการทำการทดลอง
8. สรุปผลการทดลอง

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ฟังก์ชันสูญเสียแบบใหม่ที่มีประสิทธิภาพมากกว่าฟังก์ชันสูญเสียที่มีอยู่

บทที่ 2
การทบทวนวรรณกรรมที่เกี่ยวข้อง

2.1 ปัญหาความไม่สมดุลกันของกลุ่มข้อมูล

ความไม่สมดุลกันของกลุ่มข้อมูล คือ การที่ตัวอย่างของข้อมูลแต่ละกลุ่มมีจำนวนไม่เท่ากัน และจำนวนตัวอย่างนั้นต่างกันมาก เช่น ชุดข้อมูล A มี 2 กลุ่มข้อมูลจากทั้งหมด 10,500 ตัวอย่าง แบ่งออกเป็นกลุ่มข้อมูลที่ 1 จำนวน 500 ตัวอย่าง และกลุ่มข้อมูลที่ 2 จำนวน 10,000 เป็นต้น

ในงานวิจัย[3] ผู้วิจัยได้ทำการศึกษาเกี่ยวกับผลกระทบของความไม่สมดุลกันของข้อมูลในการเรียนรู้ของแบบจำลอง และพบว่าความไม่สมดุลกันของข้อมูลได้ส่งผลกระทบด้านลบต่อกระบวนการ Backpropagation โดยผลกระทบดังกล่าว คือ การที่กลุ่มข้อมูลส่วนมากมีอิทธิพลต่อค่า Gradient ที่จะถูกนำไปใช้ในการปรับค่า Weight มากกว่ากลุ่มข้อมูลส่วนน้อย เนื่องจากจำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากในแต่ละ Batch ของการเรียนรู้ นั้นมีมากกว่า ทำให้ค่าสูญเสียรวมมีลักษณะที่ค่าสูญของกลุ่มข้อมูลส่วนมากไปกลบค่าสูญเสียของกลุ่มข้อมูลส่วนน้อย

เหตุการณ์ดังกล่าวทำให้ลักษณะของการเรียนรู้ของแบบจำลองมุ่งไปที่การเรียนรู้เฉพาะกลุ่มข้อมูลส่วนมาก กล่าวคือ ค่าสูญเสียของกลุ่มข้อมูลส่วนมากจะลดลงอย่างรวดเร็ว ในขณะที่ค่าสูญเสียของกลุ่มข้อมูลส่วนน้อยเพิ่มขึ้นเรื่อย ๆ ในช่วงต้นของการเรียนรู้ สุดท้ายทำให้การเรียนรู้ของแบบจำลองเข้าลู่เข้าจุดที่ดีที่สุดช้าหรือไม่สามารถเรียนรู้ที่จะจัดกลุ่มได้เลย

ความไม่สมดุลกันของกลุ่มข้อมูลนั้นมีอยู่ 2 ประเภท คือ Step Imbalance[1] และ Long-Tailed Imbalance[4] ตามรายละเอียดดังนี้

- Step Imbalance เป็นความไม่สมดุลกันของกลุ่มข้อมูลที่จำนวนตัวอย่างของกลุ่มข้อมูลส่วนน้อยแต่ละกลุ่มมีจำนวนเท่ากัน และจำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากแต่ละกลุ่มมีจำนวนเท่ากัน โดยอัตราส่วนของกลุ่มของส่วนน้อยและส่วนมาก (μ) สามารถคำนวณได้จากสมการที่ 2.1 ตัวอย่างการกระจายของจำนวนตัวอย่างของกลุ่มข้อมูลแสดงดังรูปที่ 2.1(ก) และ 2.1(ข)

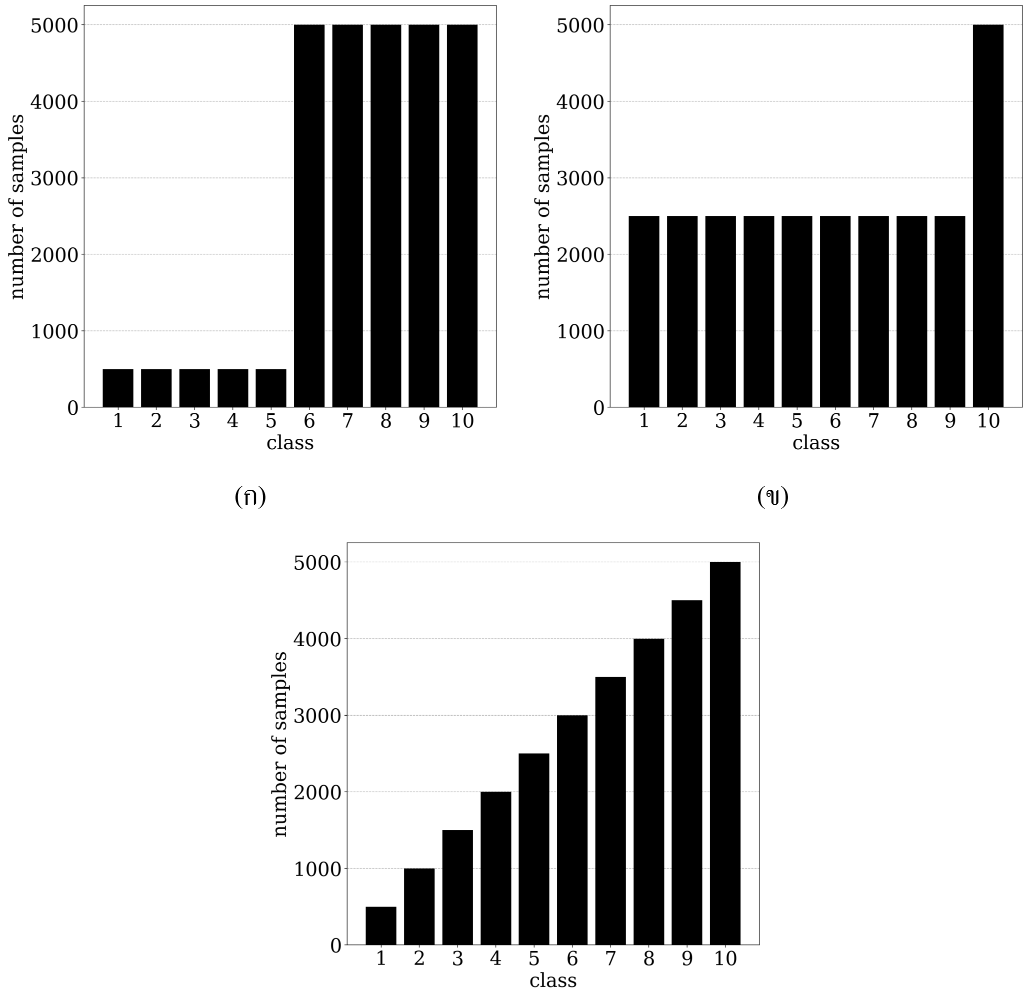
$$\mu = \frac{|\{i \in \{1, ..., N\} : C_i \; is \; minority \; class\}|}{N}, \tag{2.1}$$

  โดยที่ C_i คือ ชุดของตัวอย่างของกลุ่มข้อมูล i และ N คือ จำนวนกลุ่มตัวอย่างทั้งหมด

- Long-Tailed Imbalance เป็นความไม่สมดุลกันของกลุ่มข้อมูลที่จำนวนตัวอย่างของแต่ละกลุ่มข้อมูลมีจำนวนไม่เท่ากันตามตัวอย่างการกระจายของจำนวนตัวอย่างของกลุ่มข้อมูลในรูปที่ 2.1(ค)

สามารถคำนวณค่าอัตราส่วนระหว่างจำนวนตัวอย่างของกลุ่มข้อมูลส่วนน้อยและกลุ่มข้อมูลส่วนมาก (p) ได้ตามสมการที่ 2.2

$$p = \frac{max_i\,\{|C_i|\}}{min_i\,\{|C_i|\}} \tag{2.2}$$



(ก) (ข) (ค)

รูปที่ 2.1: ตัวอย่างการกระจายของจำนวนตัวอย่างของกลุ่มข้อมูล (ก) $p = 10$, $\mu = 0.5$ (ข) $p = 2$, $\mu = 0.9$ (ค) $p = 10$

เพื่อที่จะพิสูจน์ว่าวิธีการที่นำเสนอในงานวิจัยนี้นั้นมีประสิทธิภาพ ในเบื้องต้นจะมุ่งศึกษาที่การจัดกลุ่มข้อมูลสองกลุ่ม ที่ซึ่งลักษณะของความไม่สมดุลกันจะเป็นแบบ Step Imbalance และในอนาคตจะทำการพิสูจน์วิธีการที่นำเสนอกับการจัดกลุ่มข้อมูลหลายกลุ่ม โดยในรายงานฉบับนี้จะขอกล่าวถึงกลุ่มข้อมูลส่วนมากว่า Negative Class และกลุ่มข้อมูลส่วนน้อยว่า Positive Class

2.2 ฟังก์ชันสูญเสีย (Loss Function)

ในกระบวนการเรียนรู้ของแบบจำลอง DNN ฟังก์ชันสูญเสียจะถูกใช้ในการคำนวณค่าสูญเสีย เพื่อนำไปปรับค่า Weight ของแบบจำลอง เมื่อค่าสูญเสียยิ่งมาก ค่า Weight จะถูกปรับจากค่าเดิมมาก ในทางเดียวกันถ้าค่าสูญเสียน้อย ค่า Weight จะถูกปรับจากค่าเดิมน้อยเช่นกัน หรือก็คือแบบจำลองเริ่มไม่เรียนรู้อะไรเพิ่มเติมแล้ว ดังนั้นค่าสูญเสียต้องเป็นค่าที่เหมาะสมให้มากที่สุด ไม่เช่นนั้นจะทำให้การปรับค่า Weight ของแบบจำลองเกิดการคลาดเคลื่อนได้ อย่างไรก็ตามปัญหานี้จะพบได้ในเฉพาะการเรียนรู้ของแบบจำลองกับข้อมูลที่ไม่สมดุล ดังตัวอย่างการคำนวณค่าสูญเสียใน [5] โดยที่จำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากและส่วนน้อยเท่ากับ 90 และ 10 ตัวอย่างตามลำดับ และแบบจำลองทำนายข้อมูลของกลุ่มข้อมูลส่วนมากผิดไป 4 ตัวอย่าง และสำหรับกลุ่มข้อมูลส่วนน้อยผิดไป 5 ตัวอย่าง ถ้าคำนวณค่าสูญเสียด้วย Mean Squared Error (MSE) จะได้ค่าสูญเสียเท่ากับ 0.09 ซึ่งจากค่าสูญเสียดังกล่าว มันไม่สมเหตุสมผลเลยที่ค่าสูญเสียจะน้อยขนาดนี้ เพราะแบบจำลองทำนายข้อมูลของกลุ่มข้อมูลส่วนน้อยผิดไปตั้งครึ่งหนึ่ง ด้วยเหตุนี้ทำให้ในการเรียนรู้ของแบบจำลองในรอบถัดไป ถูกกำหนดให้เปลี่ยนแปลงค่า Weight ไม่มาก ดังนั้นจากปัญหาที่กล่าวมาจำเป็นจะต้องมีฟังก์ชันสูญเสียที่สามารถคำนวณค่าสูญเสียได้อย่างสมเหตุสมผลที่สุด เพื่อให้การเรียนรู้ของแบบจำลองนั้นมีประสิทธิภาพ

บทที่ 3
วิธีการดำเนินการวิจัย

3.1 Hybrid Loss: ฟังก์ชันสูญเสียที่นำเสนอ

3.1.1 Focal Loss

สิ่งดลใจในการคิดค้น Focal Loss (FL)[6] คือ Cross Entropy ไม่สามารถควบคุมความสมดุลระหว่างค่าสูญเสียจากกลุ่มข้อมูลส่วนมากและค่าสูญเสียจากกลุ่มข้อมูลส่วนน้อยได้ เพราะข้อมูลจากทั้งสองกลุ่มไม่สมดุลกัน แม้ว่าการใส่ Weighting Factor (α) จะสามารถแก้ปัญหานี้ได้ในเบื้องต้น แต่มันก็ไม่สามารถที่จะแยกความแตกต่างระหว่าง ตัวอย่างที่ง่าย (ตัวอย่างที่ให้ค่าสูญเสียน้อย) และ ตัวอย่างที่ยาก (ตัวอย่างที่ให้ค่าสูญเสียมาก) ได้ ซึ่งตัวอย่างที่ง่ายของกลุ่มข้อมูลส่วนมากจะส่วนในการคำนวณค่าสูญเสียรวมมาก ทำให้มีอิทธิพลต่อการคำนวณค่า Gradient มากเช่นเดียวกัน ซึ่งโดยปกติแล้วตัวอย่างที่ยากของกลุ่มข้อมูลส่วนมากจะประกอบไปด้วยข้อมูลที่เป็นประโยชน์ต่อการจัดกลุ่มมากกว่าตัวอย่างที่ง่ายของ กลุ่มข้อมูลส่วนน้อย[7] ดังนั้นการเรียนรู้จากตัวอย่างที่ยากของกลุ่มข้อมูลส่วนมากจะมีประสิทธิภาพมากกว่าการเรียนรู้จากตัวอย่างที่ง่ายของกลุ่มข้อมูลส่วนมาก

จากที่กล่าวมาข้างต้นทำให้มีความจำเป็นที่จะต้องลดการมีส่วนร่วมในการคำนวณค่าสูยเสียรวมของตัวอย่างที่ง่าย และให้ความสนใจที่การมีส่วนร่วมในการคำนวณค่าสูยเสียรวมของตัวอย่างที่ยาก ดังนั้น FL ถูกออกแบบมาเพื่อการนี้โดยการเพิ่ม Modulating Factor ($(1-p_t)^\gamma$) เข้าไปใน Cross Entropy เพื่อที่จะลดน้ำหนักของการเรียนรู้จากตัวอย่างที่ง่าย ที่ซึ่ง Modulating Factor จะช่วยลดการมีส่วนร่วมในการคำนวณค่าสูญเสียรวมจากตัวอย่างที่ง่ายและมุ่งไปที่การเรียนรู้จากตัวอย่างที่ยาก สำหรับการคำนวณค่าสูญเสียและค่าสูญเสียรวมของ FL สามารถคำนวณได้ตามสมการที่ 3.1 และ 3.2 ตามลำดับ

$$FL(p_t) = -\alpha_i(1-p_t)^\gamma \log(p_t) \tag{3.1}$$

$$l_{FL} = \frac{1}{n}\sum_{i=1}^{n} -\alpha_i(1-p_t^i)^\gamma \log(p_t^i) \tag{3.2}$$

สมการด้านบนเป็น FL ในรูปแบบที่เพิ่ม Weighting Factor เข้ามาด้วยเพื่อควบคุมความสมดุลระหว่างค่าสูญเสียจากกลุ่มข้อมูลส่วนมากและค่าสูญเสียจากกลุ่มข้อมูลส่วนน้อย โดยที่ γ คือ Focusing Parameter และสำหรับค่าของ p_t สามารถถูกระบุได้ดังสมการที่ 3.3

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{otherwise,} \end{cases} \tag{3.3}$$

โดยที่ p คือ ค่าความน่าจะเป็นในการทำนายของแบบจำลอง ดังนั้น p_t^i คือ ค่า p_t ของตัวอย่าง i

ในทางปฏิบัติ α_i จะมีค่าเท่ากับ α ถ้ากลุ่มข้อมูลของตัวอย่าง i คือ กลุ่มข้อมูลส่วนน้อย และจะมีค่าเท่ากับ $1 - \alpha$ ถ้ากลุ่มข้อมูลของตัวอย่าง i คือ กลุ่มข้อมูลส่วนมาก

3.1.2 Mean False Error

Mean False Error (MFE)[5] เป็นฟังก์ชันสูญเสียที่ถูกแก้ไขมาจาก Mean Squared Error (MSE) สิ่งจูงใจในการคิดค้นฟังก์ชันสูญเสียนี้ขึ้นมา คือ MSE ไม่สามารถตรวจจับค่าสูญเสียจากกลุ่มข้อมูลส่วนน้อยได้อย่างมีประสิทธิภาพ พูดง่าย ๆ คือ ค่าสูญเสียรวมที่คำนวณด้วย MSE มันจะมาจากค่าเฉลี่ยของค่าสูญเสียของข้อมูลทั้งหมด โดยสมการคำนวณค่าสูญเสียของ MSE และ สมการคำนวณค่าสูญเสียรวม แสดงดังสมการที่ 3.4 และ 3.5 ตามลำดับ

$$MSE = \frac{1}{2}(y - d)^2 \tag{3.4}$$

$$l_{MSE} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}(y^i - d^i)^2 \tag{3.5}$$

จากสมการด้านบน l คือ ค่าสูญเสียรวม, n คือ จำนวนตัวอย่างทั้งหมด, y^i ค่ากลุ่มข้อมูลจริงของตัวอย่าง i และ d^i คือ ค่าทำนายของแบบจำลองของตัวอย่าง i โดยที่ d สามารถคำนวณได้จากฟังก์ชัน Logistic ใด ๆ เช่น ฟังก์ชัน Sigmoid ดังสมการที่ 3.6 เป็นต้น

$$d = \frac{1}{1 + e^{-x}} \tag{3.6}$$

โดยที่ x คือ เอ๊าต์พุตจาก Layer ก่อนหน้า

จากคำนวณค่าสูญเสียรวมด้วย MSE นั้นหมายความว่าค่าสูญเสียจากกลุ่มข้อมูลส่วนมากจะไปกลบค่าสูญเสียจากกลุ่มข้อมูลส่วนน้อย เนื่องจากจำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากนั้นมีมากกว่า เพื่อที่จะแก้ปัญหานี้ MFE ถูกออกแบบให้สามารถคำนวณค่าสูญเสียรวมด้วยการรวมกันระหว่าง ค่าสูญเสียเฉลี่ยของกลุ่มข้อมูลส่วนน้อย กับ ค่าสูญเสียเฉลี่ยของกลุ่มข้อมูลส่วนมาก ตามสมการที่ 3.7 ซึ่ง

$$l_{MFE} = \frac{1}{n\_major}\sum_{i=1}^{n\_major}\frac{1}{2}(y^i - d^i)^2 + \frac{1}{n\_minor}\sum_{i=1}^{n\_minor}\frac{1}{2}(y^i - d^i)^2 \tag{3.7}$$

โดยที่ n_major คือ จำนวนตัวอย่างของกลุ่มข้อมูลส่วนมาก และ n_minor คือ จำนวนตัวอย่างของกลุ่มข้อมูลส่วนน้อย

ผลการทดลองใช้ MFE ในการเรียนรู้ของแบบจำลองกับชุดข้อมูล CIFAR-100 แสดงให้เห็นว่า MFE สามารถให้ผลการทำนายที่แม่นยำมากกว่า MSE อย่างสิ้นเชิง[5] ทั้งนี้ในงานวิจัย

ที่คิดค้น MFE ยังได้นำเสนอ Mean Squared False Error (MSFE) เพื่อที่จะเพิ่มประสิทธิภาพของ MFE โดยสามารถคำนวณได้ดังสมการที่ 3.8

$$l_{MSFE} = [\frac{1}{n\_major}\sum_{i=1}^{n\_major}\frac{1}{2}(y^i - d^i)^2]^2 + [\frac{1}{n\_minor}\sum_{i=1}^{n\_minor}\frac{1}{2}(y^i - d^i)^2]^2 \tag{3.8}$$

3.1.3 นิยามของ Hybrid Loss

คุณสมบัติที่สำคัญของ FL คือ มันสามารถควบความแตกต่างระหว่างตัวอย่างที่ง่ายและตัวอย่างที่ยาก ยิ่งไปกว่านั้นการที่เพิ่ม α เข้าไปในการคำนวณค่าสูญเสียยังช่วยทำให้ความสำคัญของตัวอย่างของกลุ่มข้อมูลส่วนน้อยและส่วนมากมีความสมดุลกัน อย่างไรก็ตามในเมื่อค่าสูญเสียรวมคือค่าเฉลี่ยของค่าสูญเสียของข้อมูลทั้งหมด มันก็ยังมีโอกาสที่ค่าสูญเสียจากกลุ่มข้อมูลส่วนมากจะไปกลบค่าสูญเสียจากกลุ่มข้อมูลส่วนน้อย ถ้าเราแก้ปัญหาด้วยการกำหนดค่า α ให้มีค่าที่มาก ก็อาจจะช่วยแก้ปัญหานี้ได้ในเบื้องต้น ผลการทดลองใน[6] ได้แสดงให้เห็นว่าการที่ α มีค่าที่มากก็ไม่ทำให้ประสทธิภาพของแบบจำลองดีไปกว่าค่าที่น้อยกว่าเลย ดังนั้นการเพิ่ม α เข้าไปในการคำนวณค่าสูญเสียอาจจะไม่เพียงพอในการแก้ปัญหานี้

เพื่อที่จะควบคุมค่าสูญเสียจากกลุ่มข้อมูลส่วนมากและกลุ่มข้อมูลส่วนน้อยให้มีความสมดุลกันอย่างสิ้นเชิง ในงานวิจัยนี้จึงได้เสนอที่จะนำรูปแบบการคำนวณค่าสูญเสียรวมของ MFE มาประยุกต์ใช้ร่วมกับการคำนวณค่าสูญเสียของ FL กล่าวคือ การคำนวณค่าสูญเสียของแต่ละตัวอย่างยังคงเหมือนเดิมตามการคำนวณด้วย FL แต่ในการคำนวณค่าสูญเสียรวมจะเป็นการนำค่าสูญเสียเฉลี่ยของกลุ่มข้อมูลส่วนมากและค่าสูญเสียเฉลี่ยของกลุ่มข้อมูลส่วนน้อยมาบวกกัน ดังสมการที่ 3.9

$$l_{Hybrid} = \frac{1}{n\_major}\sum_{i=1}^{n\_major} -\alpha_i(1-p_t^i)^{\gamma}\log(p_t^i) + \frac{1}{n\_minor}\sum_{i=1}^{n\_minor} -\alpha_i(1-p_t^i)^{\gamma}\log(p_t^i) \tag{3.9}$$

3.1.4 การหาอนุพันธ์

เพื่อที่จะสามารถนำฟังก์ชันสูญเสียไปใช้ในการเรียนรู้ของแบบจำลอง DNN ได้นั้น ฟังก์ชันสูญเสียจำเป็นต้องหาอนุพันธ์ได้ ซึ่งการหาอนุพันธ์ของแต่ละฟังก์ชันสูญเสียก็จะแตกต่างกันออกไปตามที่จะอธิบายดังต่อไปนี้

อนุพันธ์ของ FL

กำหนดให้มี x และ y โดยที่ y คือ ค่ากลุ่มข้อมูลจริง สมมติให้

$$p = \sigma(x) = \frac{1}{1+e^{-x}} \tag{3.10}$$

ดังนั้น จากสมการที่ 3.3 จะได้

$$p_t = \frac{1}{1+e^{xy}} \tag{3.11}$$

เมื่อหาอนุพันธ์ของ p_t เทียบกับ x จะได้

$$\frac{\partial p_t}{\partial x} = y(1-p_t)p_t \tag{3.12}$$

กำหนดให้

$$l_{FL} = \frac{1}{n}\sum_{i=1}^{n} -(1-p_t^i)^{\gamma}\log(p_t^i) \tag{3.13}$$

โดยที่ γ เป็นค่าคงที่ เมื่อหาอนุพันธ์ของ l_{FL} เทียบกับ x จะได้

$$\begin{aligned}\frac{\partial l_{FL}}{\partial x^i} &= \frac{1}{n}\sum_{i=1}^{n}(\frac{\partial l_{FL}}{\partial p_t^i} * \frac{\partial p_t^i}{\partial x^i})\\ &= \frac{1}{n}\sum_{i=1}^{n}[\gamma(1-p_t^i)^{\gamma-1}\log(p_t^i) + \frac{(1-p_t^i)^{\gamma}}{p_t^i}] * [y^i(1-p_t^i)p_t^i]\\ &= \frac{1}{n}\sum_{i=1}^{n} y^i(1-p_t^i)^{\gamma}(\gamma p_t^i\log(p_t^i) + p_t^i - 1)\end{aligned} \tag{3.14}$$

อนุพันธ์ของ MFE

จากสมการที่ 3.6 สามารถหาอนุพันธ์ของ d เทียบกับ x ได้ดังนี้

$$\frac{\partial d}{\partial x} = d(1-d) \tag{3.15}$$

กำหนดให้

$$MSE_{major} = \frac{1}{n\_major}\sum_{i=1}^{n\_major}\frac{1}{2}(y^i - d^i)^2 \tag{3.16}$$

$$MSE_{minor} = \frac{1}{n\_minor}\sum_{i=1}^{n\_minor}\frac{1}{2}(y^i - d^i)^2 \tag{3.17}$$

ดังนั้น

$$l_{MFE} = MSE_{major} + MSE_{minor} \tag{3.18}$$

เมื่อหาอนุพันธ์ของ l_{MFE} เทียบกับ x จะได้

$$\frac{\partial l_{MFE}}{\partial x^i} = \frac{\partial MSE_{major}}{\partial x^i} + \frac{\partial MSE_{minor}}{\partial x^i} \tag{3.19}$$

โดยที่

$$\begin{aligned}\frac{\partial MSE_{major}}{\partial x^i} &= \frac{1}{n\_major}\sum_{i=1}^{n\_major}(\frac{\partial MSE_{major}}{\partial d^i} * \frac{\partial d^i}{\partial x^i}) \\ &= -\frac{1}{n\_major}\sum_{i=1}^{n\_major}(y^i - d^i)d^i(1-d^i)\end{aligned} \tag{3.20}$$

$$\begin{aligned}\frac{\partial MSE_{minor}}{\partial x^i} &= \frac{1}{n\_minor}\sum_{i=1}^{n\_minor}(\frac{\partial MSE_{minor}}{\partial d^i} * \frac{\partial d^i}{\partial x^i}) \\ &= -\frac{1}{n\_minor}\sum_{i=1}^{n\_minor}(y^i - d^i)d^i(1-d^i)\end{aligned} \tag{3.21}$$

เราจะใช้อนุพันธ์ที่ต่างกันสำหรับตัวอย่างของแต่ละกลุ่มข้อมูล กล่าวคือ ถ้าเป็นตัวอย่างของกลุ่มข้อมูลส่วนมาก สมการที่ 3.20 จะถูกใช้ และสมการที่ 3.21 จะถูกใช้เมื่อตัวอย่างเป็นของกลุ่มข้อมูลส่วนน้อย

อนุพันธ์ของ Hybrid Loss

กำหนดให้

$$FL_{major} = \frac{1}{n\_major}\sum_{i=1}^{n\_major} -(1-p_t^i)^\gamma \log(p_t^i) \tag{3.22}$$

$$FL_{minor} = \frac{1}{n\_minor}\sum_{i=1}^{n\_minor} -(1-p_t^i)^\gamma \log(p_t^i) \tag{3.23}$$

ดังนั้น

$$l_{Hybrid} = FL_{major} + FL_{minor} \tag{3.24}$$

เมื่อหาอนุพันธ์ของ l_{Hybrid} เทียบกับ x จะได้

$$\frac{\partial l_{Hybrid}}{\partial x^i} = \frac{\partial FL_{major}}{\partial x^i} + \frac{\partial FL_{minor}}{\partial x^i} \tag{3.25}$$

โดยที่

$$\begin{aligned}\frac{\partial FL_{major}}{\partial x^i} &= \frac{1}{n\_major}\sum_{i=1}^{n\_major}(\frac{\partial FL_{major}}{\partial p_t^i} * \frac{\partial p_t^i}{\partial x^i}) \\ &= \frac{1}{n\_major}\sum_{i=1}^{n\_major} y^i(1-p_t^i)^\gamma(\gamma p_t^i \log(p_t^i) + p_t^i - 1)\end{aligned} \tag{3.26}$$

$$
\begin{aligned}
\frac{\partial FL_{minor}}{\partial x^i} &= \frac{1}{n\_minor} \sum_{i=1}^{n\_minor} (\frac{\partial FL_{minor}}{\partial p_t^i} * \frac{\partial p_t^i}{\partial x^i}) \\
&= \frac{1}{n\_minor} \sum_{i=1}^{n\_minor} y^i (1 - p_t^i)^\gamma (\gamma p_t^i \log(p_t^i) + p_t^i - 1)
\end{aligned}
\tag{3.27}
$$

เช่นเดียวกับการใช้อนุพันธ์ของ MFE ก็คือ เราจะใช้สมการที่ 3.26 ถ้าตัวอย่างเป็นของกลุ่มข้อมูลส่วนมาก และจะใช้สมการที่ 3.27 ถ้าตัวอย่างเป็นของกลุ่มข้อมูลส่วนน้อย

บทที่ 4
การทดลอง

4.1 การแบ่งข้อมูล

ในการทดลองจะแบ่งข้อมูลออกเป็น 3 ชุด คือ ชุดข้อมูลสำหรับการเรียนรู้ (Training Set), ชุดข้อมูลสำหรับการตรวจสอบ (Validation Set) และชุดข้อมูลสำหรับการทดสอบ (Testing Set) โดยในแต่ละชุดข้อมูลจะมีความไม่สมดุลกันระหว่างกลุ่มข้อมูลส่วนมากและส่วนน้อย สำหรับ Training Set กับ Validation Set จะถูกใช้ในกระบวนการเรียนรู้ของโมเดล และ Testing Set จะถูกใช้ในการทดสอบโมเดลที่เรียนรู้มาแล้ว

4.2 ชุดข้อมูล

ในงานวิจัยนี้การทดลองถูกแบ่งออกเป็น 2 ส่วนหลัก ๆ คือ การทดลองกับชุดข้อมูลที่ไม่สมดุลสำหรับการวัดเปรียบเทียบสมรรถนะเกณฑ์มาตรฐาน (Benchmark) ตามที่ถูกนำเสนอไปใน [8] และการทดลองกับชุดข้อมูลดัดแปลง โดยชุดข้อมูล CIFAR-100 [9] จะถูกดัดแปลงให้ไม่สมดุล สำหรับรายละเอียดการดัดแปลงนั้นจะถูกอธิบายไว้ในหัวข้อที่ 4.2.1

4.2.1 ชุดข้อมูลดัดแปลง

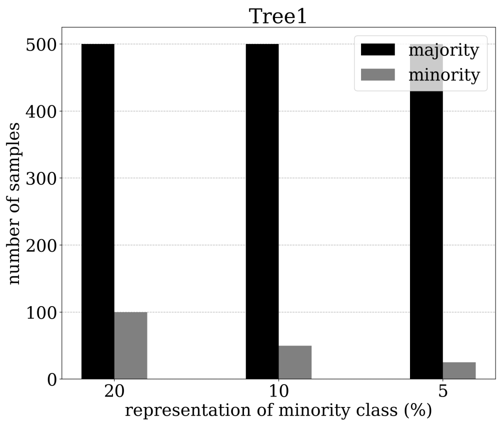
การทดลองใน [5] ได้ทำการดัดแปลง CIFAR-100 โดยการแบ่งชุดข้อมูลออกเป็น 3 ชุดข้อมูลย่อย คือ *Tree1*, *Tree2* และ *Household* และเลือกข้อมูลมา 2 กลุ่มที่แตกต่างกันสำหรับแต่ละชุดข้อมูลย่อย ที่ซึ่งข้อมูลในแต่ละชุดข้อมูลย่อยจะถูกทำให้ไม่สมดุลกัน สำหรับรายละเอียดของกลุ่มข้อมูลในแต่ละชุดข้อมูลย่อยดังนี้

Tree1 เป็นชุดข้อมูลที่ประกอบด้วยข้อมูลของกลุ่ม maple tree และ oak tree โดยจะให้กลุ่ม maple tree เป็นกลุ่มข้อมูลส่วนมาก และ oak tree เป็นกลุ่มข้อมูลส่วนน้อย

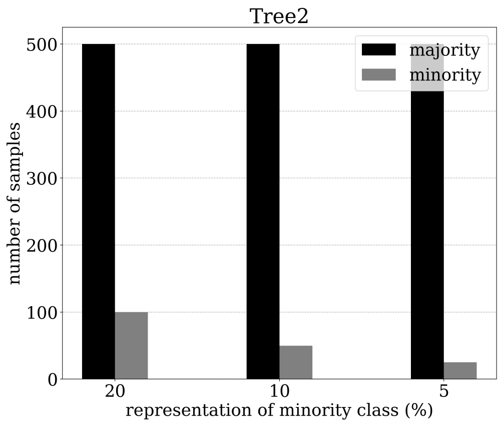
Tree2 เป็นชุดข้อมูลที่ประกอบด้วยข้อมูลของกลุ่ม maple tree และ palm tree โดยจะให้กลุ่ม maple tree เป็นกลุ่มข้อมูลส่วนมาก และ palm tree เป็นกลุ่มข้อมูลส่วนน้อย

Household เป็นชุดข้อมูลที่ประกอบด้วยข้อมูลของกลุ่ม household furniture และ household electrical devices โดยจะให้กลุ่ม household furniture เป็นกลุ่มข้อมูลส่วนมาก และ household electrical devices เป็นกลุ่มข้อมูลส่วนน้อย

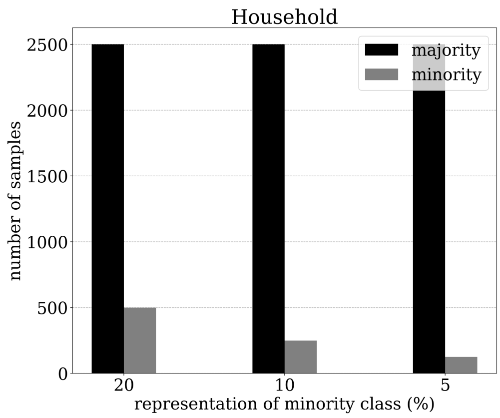
แต่ละชุดข้อมูลย่อยจะถูกแบ่งออกเป็น 3 ชุด ก็คือ ชุดข้อมูลที่มีจำนวนตัวอย่างของกลุ่มข้อมูลส่วนน้อยเป็นร้อยละ 20, 10 และ 5 ของจำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากตามลำดับ โดยจำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากและส่วนน้อยใน Training Set ของแต่ละชุดข้อมูลแสดงดังรูปที่ 4.1



(ก)



(ข)

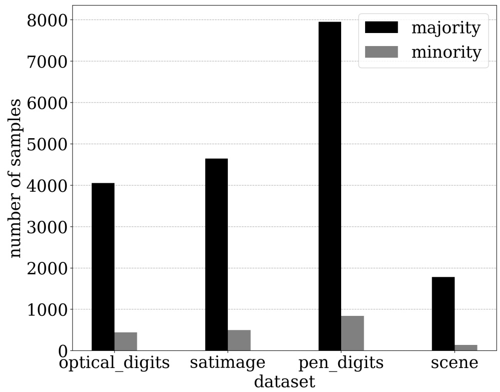


(ค)

รูปที่ 4.1: แสดงจำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากและส่วนน้อยใน Training Set ของชุดข้อมูล *Tree1*, *Tree2* และ *Household* ที่ร้อยละการมีอยู่ของตัวอย่างของกลุ่มข้อมูลส่วนน้อยต่าง ๆ

4.2.2 ชุดข้อมูล Benchmark

สำหรับชุดข้อมูล Benchmark ที่ใช้ในการทดลองประกอบไปด้วย *optical_digits*, *satimage*, *pen_digits* และ *scene* โดยสัดส่วนจำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากและส่วนน้อยใน Training ของแต่ละชุดข้อมูลแสดงดังรูปที่ 4.2



รูปที่ 4.2: แสดงจำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากและส่วนน้อยใน Training Set ของชุดข้อมูล *optical_digits* ($p = 9.14,\ \mu = 0.09$), *satimage* ($p = 9.27,\ \mu = 0.09$), *pen_digits* ($p = 9.41,\ \mu = 0.09$) และ *scene* ($p = 12.55,\ \mu = 0.07$)

4.3 การตั้งค่าเชิงเทคนิคของการทดลอง

สำหรับสถาปัตยกรรมของแบบจำลองที่ใช้ในการทดลอง คือ แบบจำลอง ResNet[10] และผู้วิจัยได้กำหนดเกณฑ์ในการหยุดกระบวนการเรียนรู้ของแบบจำลองก่อนที่จะถึงรอบสูงสุด (Early Stopping Criteria) ไว้ด้วยเพื่อความรวดเร็วในการได้มาซึ่งผลการทดลอง โดยเกณฑ์นี้จะพิจารณาที่ค่าสูญเสียของ Training Set ถ้าค่าสูญเสียไม่เปลี่ยนแปลงเป็นเวลา 10 รอบต่อเนื่อง จะทำการหยุดกระบวนการเรียนรู้ทันที สำหรับแพลตฟอร์มที่ใช้ในการพัฒนาการทดลองในงานวิจัยนี้ คือ TensorFlow[11]

4.4 Metrics สำหรับการประเมินประสิทธิภาพของแบบจำลอง

4.4.1 F1-Score

F1-Score เป็นค่าเฉลี่ยของ Precision และ Recall ที่ซึ่งเราสามารถพิจารณาประสิทธิภาพของแบบจำลองด้วย F1-Score แทนการพิจารณาด้วย Precision หรือ Recall โดย F1-Score สามารถคำนวณได้จากสมการที่ 4.1

$$F1 = 2 * \left(\frac{precision * recall}{precision + recall}\right) \tag{4.1}$$

โดยที่ Precision และ Recall สามารถคำนวณได้จากสมการที่ 4.2 และ 4.3 ตามลำดับ

$$precision = \frac{TP}{TP + FP} \tag{4.2}$$

$$recall = \frac{TP}{TP + FN} \tag{4.3}$$

โดยที่

- True Positive (TP) คือ จำนวนตัวอย่างที่แบบจำลองตรวจจับได้ว่าเป็น Positive Class และตัวอย่างเหล่านั้นเป็น Positive Class
- True Negative (TN) คือ จำนวนตัวอย่างที่แบบจำลองตรวจจับได้ว่าเป็น Negative Class และตัวอย่างเหล่านั้นเป็น Negative Class
- False Positive (FP) คือ จำนวนตัวอย่างที่แบบจำลองตรวจจับได้ว่าเป็น Positive Class แต่ตัวอย่างเหล่านั้นเป็น Negative Class
- False Negative (FN) คือ จำนวนตัวอย่างที่แบบจำลองตรวจจับได้ว่าเป็น Negative Class แต่ตัวอย่างเหล่านั้นเป็น Positive Class

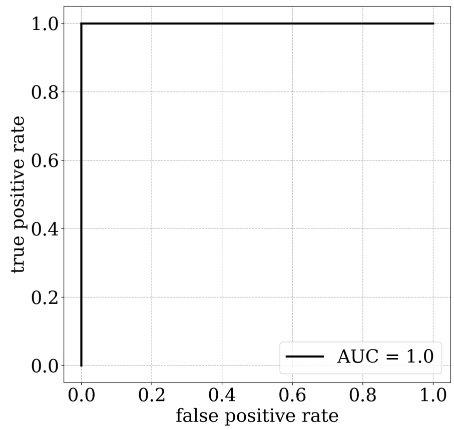
4.4.2 Area Under the ROC Curve (AUC)

AUC คือความน่าจะเป็นที่แบบจำลองจะระบุตัวอย่างของ Positive Class ว่าเป็น Positive Class และตัวอย่างของ Negative Class ว่าเป็น Negative Class โดยถ้าค่า AUC เข้าใกล้ 1 นั้นหมายความว่าแบบจำลองมีความสามารถในการแยก Positive Class ออกจาก Negative Class ได้เป็นอย่างดี ในทางเทคนิค AUC ก็คือพื้นที่ใต้กราฟของ Receiver Operating Characteristic Curve (ROC Curve) โดยความสัมพันธ์กันระหว่างค่า AUC และ ROC Curve ถูกแสดงดังรูปที่ 4.3

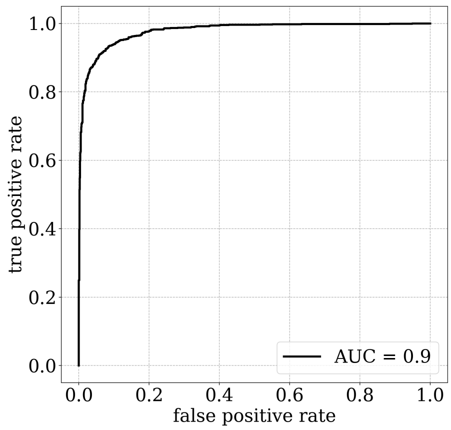
สำหรับ ROC Curve นั้นก็คือ กราฟความสัมพันธ์ระหว่าง True Positive Rate (TPR) และ False Positive Rate (FPR) โดย TPR หรือ Recall คือ ความน่าจะเป็นที่แบบจำลองสามารถตรวจจับ Positive Class จากจำนวน Positive Class ทั้งหมด และ FPR คือ ความน่าจะเป็นที่แบบจำลองจะตรวจจับ Positive Class จากจำนวน Negative Class ทั้งหมด ทั้ง TPR และ FPR สามารถคำนวณได้จากสมการที่ 4.4 และ 4.5 ตามลำดับ

$$TPR = recall = \frac{TP}{TP + FN} \tag{4.4}$$

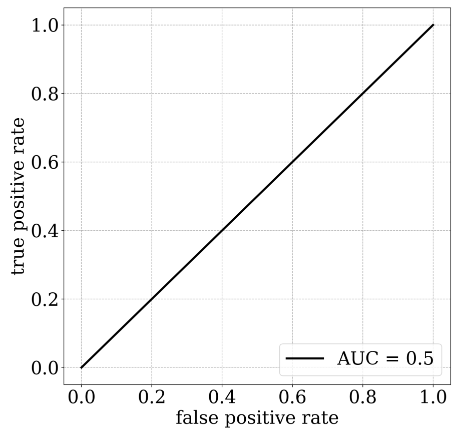
$$FPR = \frac{FP}{FP + TN} \tag{4.5}$$



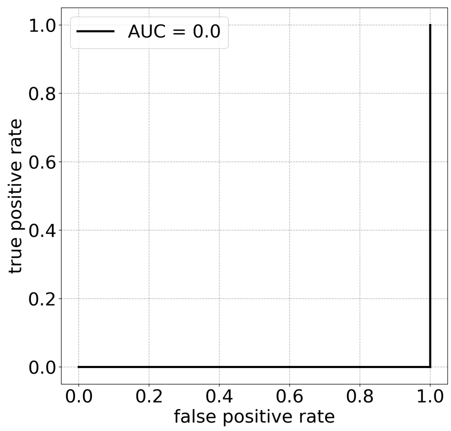
(ก)



(ข)



(ค)



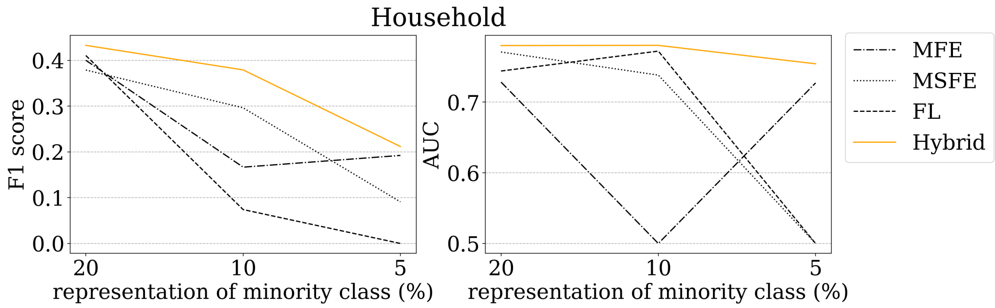
(ง)

รูปที่ 4.3: ตัวอย่างกราฟ ROC Curve แบบต่าง ๆ

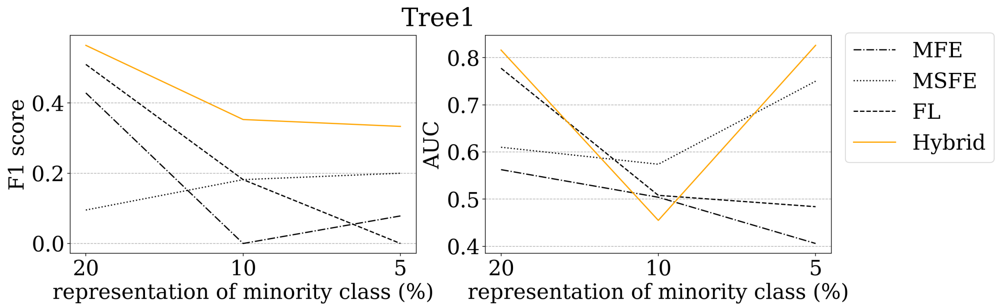
4.5 ผลการทดลอง

4.5.1 ผลการทดลองกับชุดข้อมูลดัดแปลง

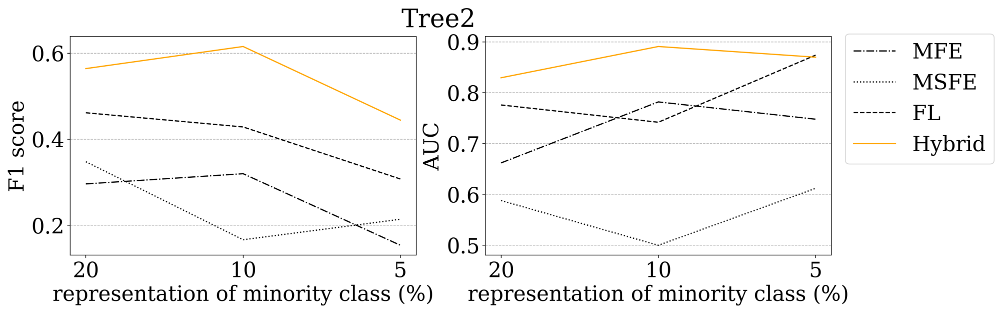
รูปที่ 4.4 แสดงผลการทดลองของแต่ละชุดข้อมูล โดยจากผลการทดลองสามารถสรุปได้ว่าฟังก์ชันสูญเสียแบบ Hybrid มีประสิทธิภาพเหนือกว่าฟังก์ชันสูญเสียอื่น ๆ อย่างสิ้นเชิง



(ก)



(ข)



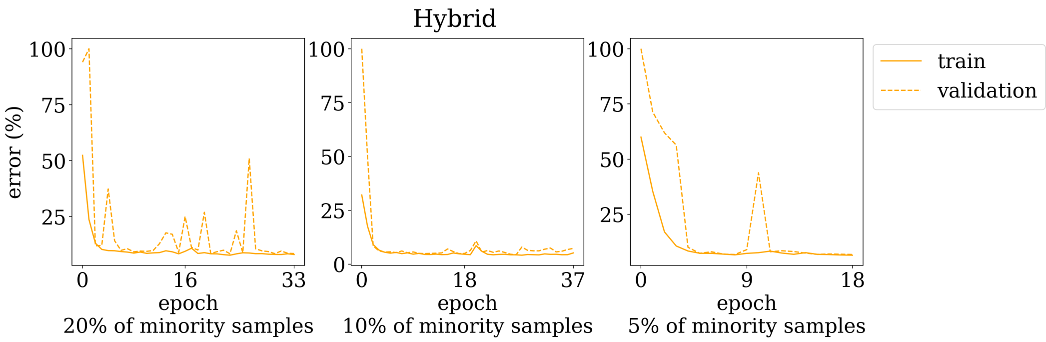
(ค)

รูปที่ 4.4: ประสิทธิภาพของแต่ละฟังก์ชันสูญเสีย เมื่อทดสอบกับชุดข้อมูลที่มีอัตราจำนวนตัวอย่างของกลุ่มข้อมูลส่วนน้อยที่แตกต่างกัน

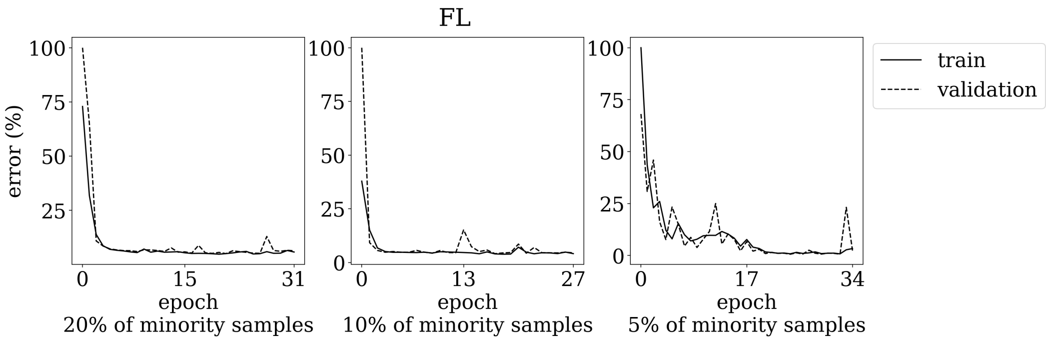
ผู้วิจัยได้สำรวจ Learning Curve ของแต่ละฟังก์ชันสูญเสียดังรูปที่ 4.5 จาก Learning Curve ของฟังก์ชันสูญเสีย Hybrid จะเห็นได้ว่าค่าสูญเสียนั้นค่อนข้างเหวี่ยงในบางชุดข้อมูล ซึ่ง Learning Curve ของฟังก์ชันสูญเสีย MSFE ก็มีลักษณะเช่นเดียวกัน อย่างไรก็ตามฟังก์ชันสูญเสีย Hybrid ยังคงมีประสิทธิภาพเชิงความแม่นยำที่ดีกว่าฟังก์ชันอื่น ๆ

จากรูปที่ 4.5 เนื่องจากในการทดลองมีการใช้ Early Stopping Criteria ด้วย จำนวนรอบในการเรียนรู้จึงต่างกัน และสามารถสรุปได้ว่าจำนวนรอบที่ใช้ในการเรียนรู้ของแต่ละฟังก์ชันสูญเสียมีความใกล้เคียงกัน

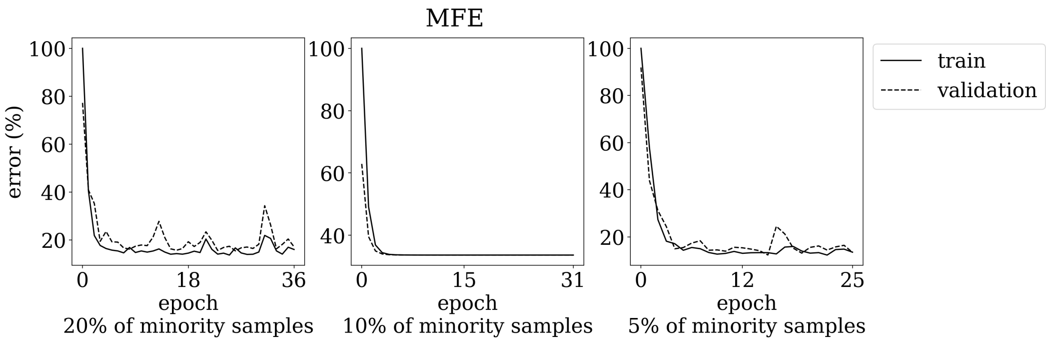
4.5.2 ผลการทดลองกับชุดข้อมูล Benchmark



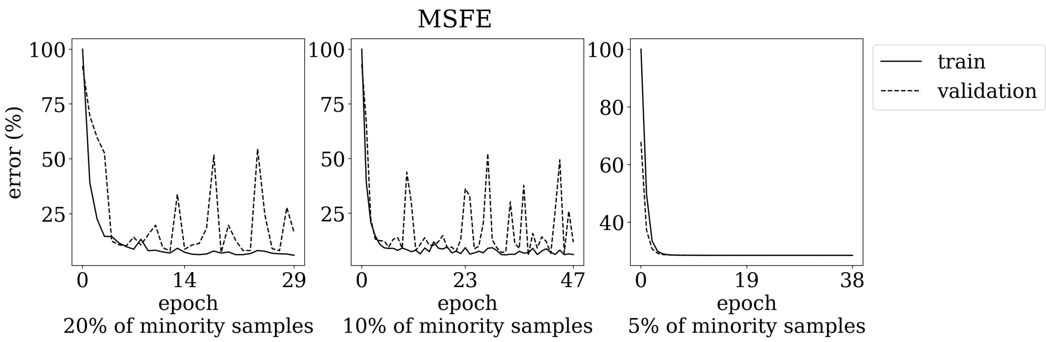
(ก)



(ข)



(ค)



(ง)

รูปที่ 4.5: ค่าสูญเสียในแต่ละรอบของการเรียนรู้ของแบบจำลอง เมื่อเรียนรู้จากชุดข้อมูล Household ที่ร้อยละการมีอยู่ของตัวอย่างของกลุ่มข้อมูลส่วนน้อยต่าง ๆ

บทที่ 5
บทสรุป

ในงายวิจัยนี้ ผู้วิจัยได้ศึกษาการทำงานของฟังก์ชันสูญเสียแบบ Mean False Error (MFE) และ Focal Loss (FL) สำหรับการเรียนรู้ของแบบจำลองโครงข่ายประสาทเทียมเชิงลึกอย่างละเอียด เพื่อที่จะหาข้อได้เปรียบของแต่ละฟังก์ชัน ฟังก์ชันสูญเสียทั้งสองดังกล่าวเป็นฟังก์ชันสูญเสียที่ถูกออกแบบมาเพื่อจัดการกับปัญหาความไม่สมดุลกันของข้อมูล จากผลการศึกษาพบว่าแต่ละฟังก์ชันสูญเสียมีวิธีการจัดการปัญหาความไม่สมดุลกันของข้อมูลที่ต่างกัน และแนวคิดของแต่ละฟังก์ชันสามารถนำมารวมกันได้ เพื่อที่จะใช้ข้อได้เปรียบของทั้งสองฟังก์ชันเพิ่มประสิทธิภาพการเรียนรู้ของโมเดล ด้วยเหตุนี้ผู้วิจัยจึงนำเสนอฟังก์ชันสูญเสียแบบใหม่ที่เรียกว่า Hybrid Loss ที่ซึ่งเป็นการผสมผสานกันระหว่าง MFE และ FL โดยการคำนวณค่าสูญเสียแต่ละตัวอย่างข้อมูลจะคำนวณด้วย FL แต่สำหรับการคำนวณค่าสูญเสียรวมจะนำวิธีการคำนวณค่าสูญเสียรวมของ MFE มาใช้ ผลการทดลองการจัดกลุ่มข้อมูลสองกลุ่มด้วยชุดข้อมูลที่หลากหลายแสดงให้เห็นว่า Hybrid Loss สามารถให้ผลการทดลองในเชิงความแม่นยำที่สูงกว่า MFE และ FL อย่างชัดเจน ซึ่งนั้นหมายความ Hybrid Loss สามารถทำงานได้อย่างดีในการจัดกลุ่มข้อมูลสองกลุ่ม ในอนาคตผู้วิจัยอาจจะทดลอง Hybrid Loss กับการจัดกลุ่มข้อมูลหลายกลุ่ม เพื่อพิสูจน์ว่าแนวคิดของ Hybrid Loss ในตอนนี้สามารถจัดการกับปัญหาความไม่สมดุลกันของข้อมูลหลายกลุ่มได้หรือไม่

บรรณานุกรม

[1] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.

[2] P. Hensman and D. Masko, "The impact of imbalanced training data for convolutional neural networks," *Degree Project in Computer Science, KTH Royal Institute of Technology*, 2015.

[3] R. Anand, K. G. Mehrotra, C. K. Mohan, and S. Ranka, "An improved algorithm for neural network classification of imbalanced training sets," *IEEE Transactions on Neural Networks*, vol. 4, no. 6, pp. 962–969, 1993.

[4] Z. Liu, Z. Miao, X. Zhan, J. Wang, B. Gong, and S. X. Yu, "Large-scale long-tailed recognition in an open world," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2537–2546.

[5] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy, "Training deep neural networks on imbalanced data sets," in *2016 international joint conference on neural networks (IJCNN)*. IEEE, 2016, pp. 4368–4374.

[6] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[7] X. Zhu, X.-Y. Jing, F. Zhang, X. Zhang, X. You, and X. Cui, "Distance learning by mining hard and easy negative samples for person re-identification," *Pattern Recognition*, vol. 95, pp. 211–222, 2019.

[8] Z. Ding, "Diversified ensemble classifiers for highly imbalanced data learning and their application in bioinformatics," 2011.

[9] A. Krizhevsky, "Learning multiple layers of features from tiny images," Tech. Rep., 2009.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[11] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.

ภาคผนวก ก
เรื่องที่หนึ่ง