

ฟังก์ชันสูญเสียสำหรับการเรียนรู้ข้อมูลที่ไม่สมดุล
Hybrid Loss for Learning Imbalanced Data

ธนวัฒน์ หลอดแก้ว
Thanawat Lodkaew
59070071

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ภาคเรียนที่ 1 ปีการศึกษา 2562

ฟังก์ชันสูญเสียสำหรับการเรียนรู้ข้อมูลที่ไม่สมดุล
Hybrid Loss for Learning Imbalanced Data

โดย

ธนวัฒน์ หลอดแก้ว
รหัสประจำตัว 59070071

อาจารย์ที่ปรึกษา
รองศาสตราจารย์ ดร. กิตติสุชาติ 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**

ใบรับรองปริญญาโท ประจำปีการศึกษา 2562
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง ฟังก์ชันสูญเสียสำหรับการเรียนรู้ข้อมูลที่ไม่สมดุล
HYBRID LOSS FOR LEARNING IMBALANCED
DATA

ผู้จัดทำ

1. ชนวัฒน์ หลอดแก้ว รหัสประจำตัว 59070071

รองศาสตราจารย์ ดร. กิติ์สุชาติ พสุภา อาจารย์ที่ปรึกษา

(.....)

ใบรับรองโครงการ (PROJECT)

เรื่อง ฟังก์ชันสูญเสียสำหรับการเรียนรู้ข้อมูลที่ไม่สมดุล Hybrid Loss for Learning Imbalanced Data

ธนวัฒน์ หลอดแก้ว รหัสประจำตัว 59070071

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

.....
ธนวัฒน์ หลอดแก้ว

ชื่อรายงาน	ฟังก์ชันสูญเสียสำหรับการเรียนรู้ข้อมูลที่ไม่สมดุล
ชื่อนักศึกษา	ธนวัฒน์ หลอดแก้ว
รหัสนักศึกษา	59070071
สาขาวิชา	เทคโนโลยีสารสนเทศ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. กิติ์สุชาติ พสุภา
ปีการศึกษา	2562

บทคัดย่อ

ปัญหาความไม่สมดุลของข้อมูล (Class Imbalance Problem) เป็นเรื่องที่ถูกหยิบขึ้นมาวิจัยอย่างแพร่หลาย เนื่องจากข้อมูลส่วนใหญ่มีจำนวนตัวอย่างของแต่ละกลุ่มข้อมูลไม่เท่ากัน ในงานวิจัยชิ้นนี้มุ่งเน้นไปที่การ สร้างสรรค์แบบจำลองการเรียนรู้เชิงลึกแบบผสม (Hybrid Deep Learning) ซึ่งเป็นการผสมผสานกันระหว่างแบบจำลองการเรียนรู้เชิงลึก (Deep Learning) หลายตัว มาเพื่อแก้ปัญหาดังกล่าว เนื่องจากแบบจำลอง Hybrid Deep Learning สามารถเพิ่มประสิทธิภาพในการจัดกลุ่มข้อมูลที่มีความสมดุลกันได้ ในทางเดียวกันแบบจำลองดังกล่าวอาจจะสามารถจัดกลุ่มข้อมูลที่ไม่สมดุลกันได้อย่างมีประสิทธิภาพเช่นเดียวกัน ในเบื้องต้นเราได้สร้างแบบจำลอง Deep Learning แบบง่าย แล้วเพิ่มความซับซ้อนของแบบจำลองขึ้นเรื่อย ๆ เพื่อที่จะได้แบบจำลองที่สมบูรณ์ที่สุด

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 classifier 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 hybrid deep learning model for classification on imbalanced data. Recently, many hybrid deep learning model can boost the accuracy on classification of balanced data. In the same way, it may work on imbalanced data effectively. Initially, a simple deep learning model is built and tested. We will find out its disadvantages or problems such as overfitting and then solve those by building more complex model. The process will be repeated until we get a perfect model with its satisfying results.

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความช่วยเหลือจาก อาจารย์ ดร. กิต์สุชาติ พสุภา อาจารย์ที่ปรึกษาปริญญานิพนธ์ ที่ได้ให้คำปรึกษาแนะนำชี้แนะแนวทางในการศึกษาค้นคว้า ตลอดจนช่วยแก้ไขข้อบกพร่องต่าง ๆ ในการทำปริญญานิพนธ์มาโดยตลอดจนโครงการนี้สำเร็จ ลุล่วงด้วยดี คณะผู้จัดทำจึงขอกราบขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณคณาจารย์ทุกท่านที่ช่วยอบรมวิชาความรู้และให้แนวคิดต่าง ๆ ที่เป็นประโยชน์และคณะเทคโนโลยีสารสนเทศสถาบันเทคโนโลยีเจ้าคุณทหารลาดกระบัง ที่เอื้อเฟื้อสถานที่ วัสดุอุปกรณ์ต่าง ๆ สำหรับจัดทำโครงการ และขอขอบคุณสื่อการสอนออนไลน์ บทความหรือวิจัยต่าง ๆ ที่ให้ข้อมูลที่เป็นประโยชน์ต่อการจัดทำโครงการครั้งนี้เป็นอย่างมาก

ขอขอบคุณ เพื่อน รุ่นพี่ภายในคณะเทคโนโลยีสารสนเทศ และผู้มีส่วนร่วมเกี่ยวข้องที่ได้ให้ คำปรึกษาและให้ความช่วยเหลือที่ดีมาตลอด สุดท้ายนี้ขอขอบพระคุณบิดามารดา และครอบครัวที่คอยให้ คำปรึกษาในเรื่องต่าง ๆ รวมทั้งเป็นกำลังใจช่วยเป็นแรงผลักดันให้ปริญญานิพนธ์นี้สำเร็จลุล่วงด้วยดี

คณะผู้จัดทำจึงขอกราบขอบพระคุณเป็นอย่างยิ่ง ไว้ ณ โอกาสนี้

ธนวัฒน์ หลอดแก้ว

ผู้จัดทำรายงาน

วันที่ 10 พฤศจิกายน พ.ศ. 2561

สารบัญ

	หน้า
บทคัดย่อ	i
บทคัดย่อ ภาษาอังกฤษ	ii
กิตติกรรมประกาศ	iii
สารบัญ	iv
สารบัญตาราง	v
สารบัญภาพ	vi
บทที่ 1 วิธีการดำเนินการวิจัย	1
1.1 Hybrid Loss	1
1.2 Metrics สำหรับการประเมินประสิทธิภาพของโมเดล	5
บทที่ 2 ผลการทดลอง	8
2.1 Binary Classification	8
2.2 Multi-Class Classification	9
บรรณานุกรม	10
ภาคผนวก ก เรื่องที่หนึ่ง	11

สารบัญตาราง

หน้า

สารบัญภาพ

	หน้า
รูปที่ 1.1 ตัวอย่างกราฟ ROC Curve แบบต่าง ๆ	7
รูปที่ 2.1 ประสิทธิภาพของแต่ละฟังก์ชันสูญเสีย เมื่อทดสอบกับชุดข้อมูลที่มีอัตรา จำนวนตัวอย่างของกลุ่มข้อมูลส่วนน้อยที่แตกต่างกัน	9

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญ

ในการเรียนรู้ของเครื่องจักร (Machine Learning) ความไม่สมดุลกันของข้อมูล หมายถึง การที่จำนวนตัวอย่างของแต่ละกลุ่มข้อมูลมีจำนวนไม่เท่ากัน ซึ่งความไม่สมดุลกันของข้อมูลนี้ถูกนิยามให้เป็นปัญหาในการจัดกลุ่มข้อมูล (Classification) สาเหตุที่ความไม่สมดุลกันของข้อมูลเป็นปัญหา คือ อัลกอริทึมการจัดกลุ่ม (Classification Algorithm) จะทำงานได้อย่างมีประสิทธิภาพก็ต่อเมื่อจำนวนตัวอย่างของแต่ละกลุ่มข้อมูลมีจำนวนที่เท่าหรือใกล้เคียงกัน เมื่อมีความไม่สมดุลกันของข้อมูลจะทำให้การทำงานของอัลกอริทึมมีประสิทธิภาพลดลง ซึ่งอาจจะลดลงจนไม่สามารถจัดกลุ่มข้อมูลได้เลย

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

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

ทำ Thresholding) การเรียนรู้แบบความเสียหายที่รู้สึกลำบาก (Cost-Sensitive Learning) การจัดกลุ่มข้อมูลแบบหนึ่งกลุ่ม (One-Class Classification) และการผนวกกันของหลายเทคนิค อย่างไรก็ตามเทคนิคเหล่านี้มียังไม่สามารถแก้ปัญหาได้อย่างมีประสิทธิภาพในทุก ๆ ชุดข้อมูล กล่าวคือ เทคนิคสามารถให้ความแม่นยำในการจัดกลุ่มข้อมูลแต่ละกลุ่มได้อย่างน่าพอใจสำหรับชุดข้อมูล A แต่ไม่สามารถทำได้มีประสิทธิภาพสำหรับชุดข้อมูล B เป็นต้น ดังนั้นเทคนิคใหม่ที่จะสามารถการแก้ปัญหาการจัดกลุ่มข้อมูลที่ไม่สมดุลกันได้อย่างมีประสิทธิภาพ และปรับเข้าได้กับทุกชุดข้อมูลจำเป็นต้องถูกคิดค้นขึ้น

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาการจัดการปัญหาความไม่สมดุลกันของข้อมูลด้วยเทคนิคต่าง ๆ

1.3 ขอบเขตการพัฒนาโครงการ

1. ทดลองใช้เทคนิคการจัดกลุ่มข้อมูลที่ไม่สมดุลกันแบบต่าง ๆ ในแต่ละชุดข้อมูล เพื่อค้นหาว่าเทคนิคที่ดีที่สุด

1.4 ขั้นตอนการดำเนินงาน

1. ศึกษาเกี่ยวกับนิยามของความไม่สมดุลกันของข้อมูลในด้านการจัดกลุ่มข้อมูล
2. ศึกษารูปแบบการแก้ปัญหการจัดกลุ่มข้อมูลที่ไม่สมดุลกันแบบต่าง ๆ
3. ศึกษางานวิจัยที่เกี่ยวข้อง
4. ตั้งข้อสมมติฐาน
5. ออกแบบการทดลอง
6. เลือกชุดข้อมูล และ Metrics ที่จะใช้ในการทดลอง
7. ดำเนินการทำการทดลอง
8. สรุปผลการทดลอง

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. เทคนิคในการ สร้างแบบจำลองการจัดกลุ่มข้อมูลที่สามารถให้ความแม่นยำที่สูงในการระบุข้อมูลแต่ละกลุ่ม แม้ว่าแบบจำลองจะเรียนรู้จากชุดข้อมูลที่ไม่สมดุลกัน

บทที่ 2

การทบทวนวรรณกรรมที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 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^[2] ผู้วิจัยได้ทำการศึกษาเกี่ยวกับผลกระทบของความไม่สมดุลกันของข้อมูลในการเรียนรู้ของโมเดล และพบว่าความไม่สมดุลกันของข้อมูลได้ส่งผลกระทบต่อกระบวนการ Backpropagation โดยผลกระทบดังกล่าว คือ การที่กลุ่มข้อมูลส่วนมากมีอิทธิพลต่อค่า Gradient ที่จะถูกนำไปใช้ในการปรับค่า Weight มากกว่ากลุ่มข้อมูลส่วนน้อย เนื่องจากจำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากในแต่ละ Batch ของการเรียนรู้ นั้นมีมากกว่า ทำให้ค่าสูญเสียรวมมีลักษณะที่ค่าสูญเสียของกลุ่มข้อมูลส่วนมากไปกลบค่าสูญเสียของกลุ่มข้อมูลส่วนน้อย

เหตุการณ์ดังกล่าวทำให้ลักษณะของการเรียนรู้ของโมเดลมุ่งไปที่การเรียนรู้เฉพาะกลุ่มข้อมูลส่วนมาก กล่าวคือ ค่าสูญเสียของกลุ่มข้อมูลส่วนมากจะลดลงอย่างรวดเร็ว ในขณะที่ค่าสูญเสียของกลุ่มข้อมูลส่วนน้อยเพิ่มขึ้นเรื่อยๆ ในช่วงต้นของการเรียนรู้ สุดท้ายทำให้การเรียนรู้ของโมเดลเข้าสู่จุดที่ดีที่สุดช้าหรือไม่สามารถเรียนรู้ที่จะจัดกลุ่มได้เลย

ความไม่สมดุลกันของกลุ่มข้อมูลนั้นมีอยู่ 2 ประเภท คือ Step Imbalance^[1] และ Long-Tailed Imbalance^[3] ตามรายละเอียดดังนี้

- Step Imbalance เป็นความไม่สมดุลกันของกลุ่มข้อมูลที่มีจำนวนตัวอย่างของกลุ่มข้อมูลส่วนน้อยแต่ละกลุ่มมีจำนวนเท่ากัน และจำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากแต่ละกลุ่มมีจำนวนเท่ากัน โดยอัตราส่วนของกลุ่มของส่วนน้อยและส่วนมาก (μ) สามารถคำนวณได้จากสมการที่ ?? ตัวอย่างการกระจายของจำนวนตัวอย่างของกลุ่มข้อมูลแสดงดังรูปที่ ?? และ ??

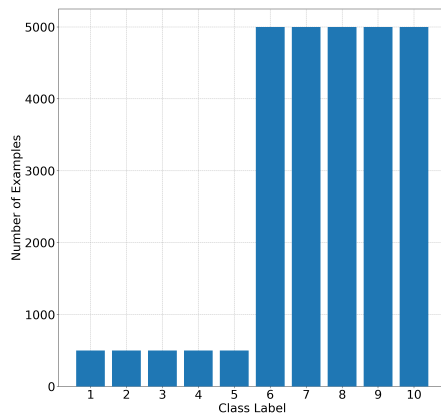
$$\mu = \frac{|\{i \in \{1, \dots, N\} : C_i \text{ is minority class}\}|}{N}, \quad (2.1)$$

โดยที่ C_i คือ ชุดของตัวอย่างของกลุ่มข้อมูล i และ N คือ จำนวนกลุ่มตัวอย่างทั้งหมด

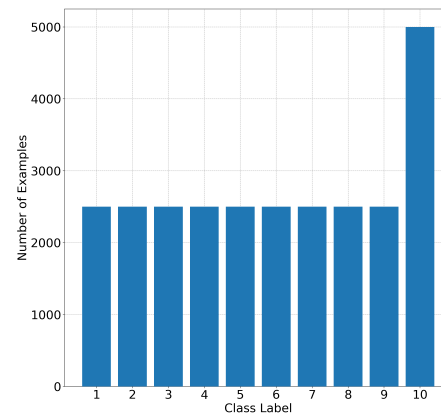
- Long-Tailed Imbalance เป็นความไม่สมดุลกันของกลุ่มข้อมูลที่มีจำนวนตัวอย่างของแต่ละกลุ่มข้อมูลมีจำนวนไม่เท่ากันตามตัวอย่างการกระจายของจำนวนตัวอย่างของกลุ่มข้อมูลในรูปที่ ??

สามารถคำนวณค่าอัตราส่วนระหว่างจำนวนตัวอย่างของกลุ่มข้อมูลส่วนน้อยและกลุ่มข้อมูลส่วนมาก (p) สำหรับความไม่สมดุลกันของกลุ่มข้อมูลทั้งสองประเภทได้ตามสมการที่ ??

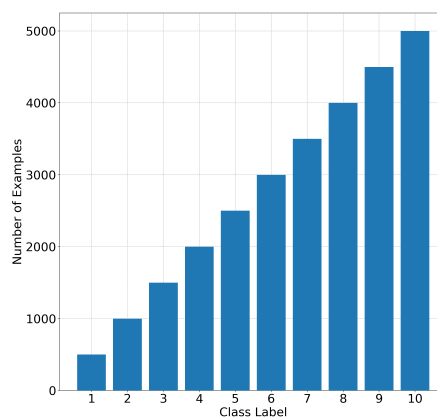
$$p = \frac{\max_i \{|C_i|\}}{\min_i \{|C_i|\}} \quad (2.2)$$



(ก)



(ข)



(ค)

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

2.1.2 การจัดการปัญหาความไม่สมดุลกันของกลุ่มข้อมูลด้วยวิธีการต่าง ๆ

Cost-Sensitive Learning

Cost-Sensitive Learning เป็นหนึ่งในเทคนิคการเรียนรู้ของอัลกอริทึมการจัดกลุ่มข้อมูลที่มีการนำ ค่าความเสียหายของการ ระบุกลุ่มข้อมูลผิดพลาด (Misclassification Cost) มา

ตารางที่ 2.1: ตัวอย่างของ Cost Matrix สำหรับการจัดกลุ่มข้อมูลสองกลุ่ม โดย 1 คือ Positive และ 0 คือ Negative

	Actual Negative	Actual Positive
Predict Negative	C(0,0), or TP	C(0,1), or FN
Predict Positive	C(1,0), or FP	C(1,1), or TP

พิจารณาเพื่อที่จะพยายามลดค่าความเสียหายดังกล่าวให้น้อยที่สุด^[4] ในกระบวนการเรียนรู้ของอัลกอริทึมการจัดกลุ่มข้อมูล ส่วนใหญ่พยายามที่จะลดอัตราความผิดพลาดในการระบุกลุ่มข้อมูลให้น้อยที่สุด ซึ่งอัตราความผิดพลาด คือ อัตราของการระบุกลุ่มข้อมูลผิด โดยแบ่งออกเป็นสองประเภท คือ ความผิดพลาดที่ระบุข้อมูลกลุ่ม Positive เป็นกลุ่ม Negative (False Negative) และ ความผิดพลาดที่ระบุข้อมูลกลุ่ม Negative เป็นกลุ่ม Positive (False Positive) อัลกอริทึมดังกล่าวจะพยายามลดความผิดพลาดโดยรวม กล่าวคือ ให้ความสำคัญที่ False Negative และ False Positive เท่ากัน

ในความเป็นจริงบางครั้งความผิดพลาดดังกล่าวไม่ควรให้ความสำคัญเท่ากัน เช่น ในการตรวจวินิจฉัยโรคมะเร็ง ที่ซึ่งผู้ป่วยที่มีโรคมะเร็งเป็นข้อมูลกลุ่ม Positive และผู้ป่วยที่มีสุขภาพดีเป็นข้อมูลกลุ่ม Negative การวินิจฉัยผิดพลาดว่าผู้ป่วยมีสุขภาพดี แต่จริง ๆ แล้วมีโรคมะเร็ง (False Negative) นั้นมีความเสียหายมากกว่าการที่วินิจฉัยผิดพลาดว่าผู้ป่วยมีโรคมะเร็ง แต่จริง ๆ แล้วมีสุขภาพดี (False Positive) ดังนั้นควรจะให้ความสำคัญที่ False Negative มากกว่าเป็นต้น

Cost-Sensitive Learning จะให้ความสำคัญกับความเสียหายดังกล่าวไม่เท่ากัน ที่ซึ่ง False Negative จะถูกให้ความสำคัญมากกว่า และเรียกว่าเป็นค่าความเสียหายของการระบุกลุ่มข้อมูลผิดพลาด ค่าความเสียหายถูกระบุไว้ดังตารางที่ ?? ซึ่งเรียกว่า Cost Matrix โดยที่ $C(i,j)$ คือ ค่าความเสียหายในการระบุข้อมูลกลุ่ม j เป็นกลุ่ม i

โดยปกติแล้ว Cost-Sensitive ถูกนำมาใช้ในการแก้ปัญหาความไม่สมดุลกันของข้อมูล เนื่องจาก โดยทั่วไปกลุ่มข้อมูลส่วนใหญ่นั้นจะเป็นกลุ่ม Positive และข้อมูลกลุ่ม Positive มักจะถูกระบุผิดว่าเป็นกลุ่ม Negative

2.1.3 การเรียนรู้เชิงลึก (Deep Learning)

Ian Goodfellow et al.^[5] ได้อธิบายว่า การเรียนรู้เชิงลึกเป็นวิธีการที่สามารถแก้ปัญหาที่ดำเนินการได้ง่ายสำหรับมนุษย์ แต่ยากต่อการอธิบาย กล่าวคือ ปัญหาเหล่านั้นเป็นปัญหาที่เราสามารถแก้ได้อย่างอัตโนมัติด้วยสัญชาตญาณ เช่น การจดจำคำพูด หรือใบหน้าผู้คนในรูป เป็นต้น สามารถเรียกปัญหาหรืองานประเภทนี้ว่า งานที่เป็นรูปแบบ (Formal Task) ในทางตรงกันข้าม ปัญหาหรืองานที่ไม่เป็นรูปแบบ (Informal Task) ที่ซึ่งยากสำหรับมนุษย์ในการดำเนินการ แต่ง่ายสำหรับการประมวลผลของคอมพิวเตอร์ ปัญหาประเภทนี้จะสามารถถูกอธิบายออกมาในรูปแบบของสมการคณิตศาสตร์ หรือมีวิธีการคำนวณทางคณิตศาสตร์ได้ เช่น การพยากรณ์สภาพอากาศ ที่ซึ่งปัจจัยต่าง ๆ ที่เกี่ยวข้อง อย่าง ค่าความชื้น ค่าอุณหภูมิวัน

ก่อนหน้า และค่าอื่น ๆ จะถูกนำมาคำนวณ เป็นต้น

การเรียนรู้เชิงลึกทำให้คอมพิวเตอร์สามารถเรียนรู้จากประสบการณ์ และเข้าใจสิ่งต่าง ๆ แบบเป็นลำดับชั้นของแนวคิดเหมือนกับมนุษย์ โดยที่แต่ละลำดับชั้นของแนวคิดถูกกำหนดผ่านความสัมพันธ์กับชั้นของแนวคิดที่ไม่ซับซ้อน กล่าวคือ แนวคิดที่เป็นลำดับชั้นทำให้คอมพิวเตอร์สามารถเรียนรู้แนวคิดที่ซับซ้อนโดยการสร้างมันจากแนวคิดที่ไม่ซับซ้อน

ในศาสตร์ทางด้านปัญญาประดิษฐ์ (Artificial Intelligence: AI) การที่ระบบคอมพิวเตอร์สามารถได้รับมาซึ่งความรู้จากการสกัดรูปแบบของข้อมูลดิบ เรียกความสามารถนี้ว่า การเรียนรู้ของเครื่องจักร อัลกอริทึมพื้นฐานของการเรียนรู้ของเครื่องจักรอย่าง Logistic Regression Linear Regression หรือ Naive Bayes สามารถประมวลเพื่อดำเนินงานที่เป็นรูปแบบต่าง ๆ เช่น พยากรณ์สภาพอากาศ แนะนำช่วงเวลาสำหรับการคลอดลูก หรือแยกแยะอีเมลที่ต้องออกจากอีเมลขยะ เป็นต้น ประสิทธิภาพของอัลกอริทึมพื้นฐานเหล่านั้นขึ้นอยู่กับคุณลักษณะ (Feature) ของข้อมูลที่ได้รับ ตัวอย่างเช่น เมื่อการถดถอยโลจิสติกส์ถูกใช้ในการพยากรณ์โรคมะเร็งเต้านม हमจะต้องให้ข้อมูลที่เกี่ยวข้องแก่อัลกอริทึม เช่น เต้านมมีผื่นแดง ร้อน ผื่นคล้ายผิวส้มหรือไม่ และมีก้อนหนาๆ ในเต้านมหรือใต้แขนหรือไม่ เป็นต้น ข้อมูลดังกล่าวเป็นคุณลักษณะของคนไข้ ซึ่งเรียกแต่ละประเภทข้อมูลว่า คุณลักษณะ การถดถอยโลจิสติกส์จะเรียนรู้ว่าคุณลักษณะเหล่านั้นมีความสัมพันธ์กันอย่างไร กับผลลัพธ์ ซึ่งผลลัพธ์ในตัวอย่างนี้คือ โอกาสในการเป็นโรคมะเร็งเต้านม อย่างไรก็ตามอัลกอริทึมไม่สามารถกำหนดได้ว่าคุณลักษณะต้องถูกกำหนดมาอย่างไร กล่าวคือ การได้มาซึ่งค่าของคุณลักษณะไม่ใช่หน้าที่ของอัลกอริทึม เพราะฉะนั้นถ้าอัลกอริทึมได้รับภาพ MRI มาเพื่อพยากรณ์โรคมะเร็งเต้านม แทนการใช้ข้อมูลที่เป็นรูปแบบของหโม จะทำอัลกอริทึมไม่สามารถพยากรณ์ผลลัพธ์ที่เป็นประโยชน์ได้ หรือผลลัพธ์ที่ได้จะไม่น่าเชื่อถือเลย เพราะค่า Pixel ในภาพ MRI ไม่ได้มีความสัมพันธ์กับลักษณะความผิดปกติของเต้านมอย่างสิ้นเชิง

แบบจำลองพื้นฐานของการเรียนรู้เชิงลึก คือ Multi-Layer Perceptron (MLP) ที่ซึ่งเป็นเพียงฟังก์ชันคณิตศาสตร์ที่ทำการแปลงค่าข้อมูลขาเข้าไปเป็นข้อมูลขาออก ฟังก์ชันจะถูกสร้างขึ้นจากการ ประกอบเข้าด้วยกัน ของหลายฟังก์ชันที่เป็นพื้นฐานกว่า เราสามารถใช้ฟังก์ชันที่ต่างกันในแต่ละงานเพื่อให้ได้มาซึ่งคุณลักษณะของข้อมูลที่เหมาะสมกับงานนั้น ๆ

2.1.4 การเรียนรู้ของโมเดล Deep Learning

2.1.5 ฟังก์ชันสูญเสีย (Loss Function)

2.2 งานวิจัยที่เกี่ยวข้อง

เมื่อเทคนิคการจัดกลุ่มข้อมูลที่ประกอบด้วยโครงข่ายประสาทแบบคอนโวลูชันถูกใช้กับชุดข้อมูลที่มีความไม่สมดุลกัน อัตราการพยากรณ์ผิดพลาดจะมีค่าสูงเมื่อเทียบกับการเพิ่มขึ้นของจำนวนรอบของการเรียนรู้ของแบบจำลอง กล่าวคือ ยิ่งจำนวนรอบของการเรียนรู้สูงขึ้น จะทำให้อัตราการพยากรณ์ผิดพลาดสูงขึ้นด้วย^[6] เบื้องหลังของสาเหตุที่ทำให้เป็นเช่น

นั่น คือ ในเสตจของการเรียนรู้ของแบบจำลองการเรียนรู้เชิงลึก ข้อมูลจะถูกแบ่งออกเป็นกลุ่ม ๆ ซึ่งทำให้แต่ละกลุ่มมีความไม่เท่าเทียมกันเมื่อข้อมูลไม่สมดุลกัน อีกทั้งบางกลุ่มอาจจะมีแค่ตัวอย่างของกลุ่มข้อมูลที่เป็นส่วนมาก หรือกลุ่มข้อมูลที่เป็นส่วนน้อยเท่านั้น เมื่อแบบจำลองได้เรียนรู้ข้อมูลจากกลุ่มเหล่านั้นในทุก ๆ รอบ จึงทำให้เกิดอัตราการพยากรณ์ผิดพลาดที่สูง

เมื่อไม่นานมานี้ได้มีงานวิจัยที่พยายามจัดการกับปัญหาการเรียนรู้ของแบบจำลองการเรียนรู้เชิงลึกเมื่อต้องเรียนรู้ข้อมูลที่มีความไม่สมดุลกัน ดังต่อไปนี้

บทที่ 3

วิธีการดำเนินการวิจัย

3.1 Hybrid Loss

3.1.1 Focal Loss

สิ่งคลอใจในการคิดค้น Focal Loss (FL)^[7] คือ Cross Entropy ไม่สามารถควบคุมความสมดุลระหว่างค่าสูญเสียจากกลุ่มข้อมูลส่วนมากและค่าสูญเสียจากกลุ่มข้อมูลส่วนน้อยได้ เพราะข้อมูลจากทั้งสองกลุ่มไม่สมดุลกัน แม้ว่าการใส่ Weighting Factor (α) จะสามารถแก้ปัญหานี้ได้ในเบื้องต้น แต่มันก็ไม่สามารถที่จะแยกความแตกต่างระหว่าง ตัวอย่างที่ง่าย (ตัวอย่างที่ให้ค่าสูญเสียน้อย) และ ตัวอย่างที่ยาก (ตัวอย่างที่ให้ค่าสูญเสียมาก) ได้ ซึ่งตัวอย่างที่ง่ายของกลุ่มข้อมูลส่วนมากจะส่วนในการคำนวณค่าสูญเสียรวมมาก ทำให้มีอิทธิพลต่อการคำนวณค่า Gradient มากเช่นเดียวกัน ซึ่งโดยปกติแล้วตัวอย่างที่ยากของกลุ่มข้อมูลส่วนมากจะประกอบไปด้วยข้อมูลที่เป็นประโยชน์ต่อการจัดกลุ่มมากกว่าตัวอย่างที่ง่ายของกลุ่มข้อมูลส่วนน้อย^[8] ดังนั้นการเรียนรู้จากตัวอย่างที่ยากของกลุ่มข้อมูลส่วนมากจะมีประสิทธิภาพมากกว่าการเรียนรู้จากตัวอย่างที่ง่ายของกลุ่มข้อมูลส่วนมาก

จากที่กล่าวมาข้างต้นทำให้มีความจำเป็นที่จะต้องลดการมีส่วนร่วมในการคำนวณค่าสูญเสียรวมของตัวอย่างที่ง่าย และให้ความสนใจที่การมีส่วนร่วมในการคำนวณค่าสูญเสียรวมของตัวอย่างที่ยาก ดังนั้น FL ถูกออกแบบมาเพื่อการนี้โดยการเพิ่ม Modulating Factor $((1 - p_t)^\gamma)$ เข้าไปใน Cross Entropy เพื่อที่จะลดน้ำหนักของการเรียนรู้จากตัวอย่างที่ง่าย ที่ซึ่ง Modulating Factor จะช่วยลดการมีส่วนร่วมในการคำนวณค่าสูญเสียรวมจากตัวอย่างที่ง่าย และมุ่งไปที่การเรียนรู้จากตัวอย่างที่ยาก สำหรับการคำนวณค่าสูญเสียและค่าสูญเสียรวมของ FL สามารถคำนวณได้ตามสมการที่ 1.1 และ 1.2 ตามลำดับ

$$FL(p_t) = -\alpha_i(1 - p_t)^\gamma \log(p_t) \quad (3.1)$$

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

สมการด้านบนเป็น FL ในรูปแบบที่เพิ่ม Weighting Factor เข้ามาด้วยเพื่อควบคุมความสมดุลระหว่างค่าสูญเสียจากกลุ่มข้อมูลส่วนมากและค่าสูญเสียจากกลุ่มข้อมูลส่วนน้อย โดยที่ γ คือ Focusing Parameter และสำหรับค่าของ p_t สามารถถูกระบุได้ดังสมการที่ 1.3

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

โดยที่ p คือ ค่าความน่าจะเป็นในการทำนายของโมเดล 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)^[9] เป็นฟังก์ชันสูญเสียที่ถูกแก้ไขมาจาก Mean Squared Error (MSE) สิ่งจูงใจในการคิดค้นฟังก์ชันสูญเสียนี้ขึ้นมา คือ MSE ไม่สามารถตรวจจับค่าสูญเสียจากกลุ่มข้อมูลส่วนน้อยได้อย่างมีประสิทธิภาพ พุดง่าย ๆ คือ ค่าสูญเสียรวมที่คำนวณด้วย MSE มันจะมาจากค่าเฉลี่ยของค่าสูญเสียของข้อมูลทั้งหมด โดยสมการคำนวณค่าสูญเสียของ MSE และ สมการคำนวณค่าสูญเสียรวม แสดงดังสมการที่ 1.4 และ 1.5 ตามลำดับ

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

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

จากสมการด้านบน l คือ ค่าสูญเสียรวม, n คือ จำนวนตัวอย่างทั้งหมด, y^i ค่ากลุ่มข้อมูลจริงของตัวอย่าง i และ d^i คือ ค่าทำนายของโมเดลของตัวอย่าง i โดยที่ d สามารถคำนวณได้จากฟังก์ชัน Logistic ใด ๆ เช่น ฟังก์ชัน Sigmoid ดังสมการที่ 1.6 เป็นต้น

$$d = \frac{1}{1 + e^{-x}} \quad (3.6)$$

โดยที่ x คือ เวกเตอร์จาก Layer ก่อนหน้า

จากคำนวณค่าสูญเสียรวมด้วย MSE นั้นหมายความว่าค่าสูญเสียจากกลุ่มข้อมูลส่วนมากจะไปกลบค่าสูญเสียจากกลุ่มข้อมูลส่วนน้อย เนื่องจากจำนวนตัวอย่างของกลุ่มข้อมูลส่วนมากนั้นมีมากกว่า เพื่อที่จะแก้ปัญหานี้ MFE ถูกออกแบบให้สามารถคำนวณค่าสูญเสียรวมด้วยการรวมกันระหว่าง ค่าสูญเสียเฉลี่ยของกลุ่มข้อมูลส่วนน้อย กับ ค่าสูญเสียเฉลี่ยของกลุ่มข้อมูลส่วนมาก ตามสมการที่ 1.7 ซึ่ง

$$l_{MFE} = \frac{1}{n_{major}} \sum_{i=1}^{n_{major}} \frac{1}{2}(y^i - d^i)^2 + \frac{1}{n_{minor}} \sum_{i=1}^{n_{minor}} \frac{1}{2}(y^i - d^i)^2 \quad (3.7)$$

โดยที่ n_{major} คือ จำนวนตัวอย่างของกลุ่มข้อมูลส่วนมาก และ n_{minor} คือ จำนวนตัวอย่างของกลุ่มข้อมูลส่วนน้อย

ผลการทดลองใช้ MFE ในการเรียนรู้ของโมเดลกับชุดข้อมูล CIFAR-100 แสดงให้เห็นว่า MFE สามารถให้ผลการทำนายที่แม่นยำมากกว่า MSE อย่างสิ้นเชิง^[9]

3.1.3 นิยามของ Hybrid Loss

คุณสมบัติที่สำคัญของ FL คือ มันสามารถควบคุมความแตกต่างระหว่างตัวอย่างที่ง่ายและตัวอย่างที่ยาก ยิ่งไปกว่านั้นการที่เพิ่ม α เข้าไปในการคำนวณค่าสูญเสียยังช่วยทำให้ความ

สำคัญของตัวอย่างของกลุ่มข้อมูลส่วนน้อยและส่วนมากมีความสมดุลกัน อย่างไรก็ตามในเมื่อค่าสูญเสียรวมคือค่าเฉลี่ยของค่าสูญเสียของข้อมูลทั้งหมด มันก็ยังมีโอกาสที่ค่าสูญเสียจากกลุ่มข้อมูลส่วนมากจะไปกลบค่าสูญเสียจากกลุ่มข้อมูลส่วนน้อย ถ้าเราแก้ปัญหด้วยการกำหนดค่า α ให้มีค่าที่มาก ก็อาจจะช่วยแก้ปัญหาได้ในเบื้องต้น ผลการทดลองใน^[7] ได้แสดงให้เห็นว่าการที่ α มีค่าที่มากก็ไม่ทำให้ประสิทธิภาพของโมเดลดีไปกว่าค่าที่น้อยกว่าเลย ดังนั้นการเพิ่ม α เข้าไปในการคำนวณค่าสูญเสียอาจจะไม่เพียงพอในการแก้ปัญหา

เพื่อที่จะควบคุมค่าสูญเสียจากกลุ่มข้อมูลส่วนมากและกลุ่มข้อมูลส่วนน้อยให้มีความสมดุลกันอย่างสิ้นเชิง ในงานวิจัยนี้จึงได้เสนอที่จะนำรูปแบบการคำนวณค่าสูญเสียรวมของ MFE มาประยุกต์ใช้ร่วมกับการคำนวณค่าสูญเสียของ FL กล่าวคือ การคำนวณค่าสูญเสียของแต่ละตัวอย่างยังคงเหมือนเดิมตามการคำนวณด้วย FL แต่ในการคำนวณค่าสูญเสียรวมจะเป็นการนำค่าสูญเสียเฉลี่ยของกลุ่มข้อมูลส่วนมากและค่าสูญเสียเฉลี่ยของกลุ่มข้อมูลส่วนน้อยมาบวกกัน ดังสมการที่ 1.8

$$l_{Hybrid} = \frac{1}{n_{major}} \sum_{i=1}^{n_{major}} -\alpha_i(1 - p_t^i)^\gamma \log(p_t^i) + \frac{1}{n_{minor}} \sum_{i=1}^{n_{minor}} -\alpha_i(1 - p_t^i)^\gamma \log(p_t^i) \quad (3.8)$$

3.1.4 การหาอนุพันธ์

อนุพันธ์ของ FL

กำหนดให้มี x และ y โดยที่ y คือ ค่ากลุ่มข้อมูลจริง สมมติให้

$$p = \sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.9)$$

ดังนั้น จากสมการที่ 1.3 จะได้

$$p_t = \frac{1}{1 + e^{xy}} \quad (3.10)$$

เมื่อหาอนุพันธ์ของ p_t เทียบกับ x จะได้

$$\frac{\partial p_t}{\partial x} = y(1 - p_t)p_t \quad (3.11)$$

กำหนดให้

$$l_{FL} = \frac{1}{n} \sum_{i=1}^n -(1 - p_t^i)^\gamma \log(p_t^i) \quad (3.12)$$

โดยที่ γ เป็นค่าคงที่ เมื่อหาอนุพันธ์ของ l_{FL} เทียบกับ x จะได้

$$\begin{aligned}
\frac{\partial l_{FL}}{\partial x^i} &= \frac{\partial l_{FL}}{\partial p_t^i} * \frac{\partial p_t^i}{\partial x^i} \\
&= \frac{1}{n} \sum_{i=1}^n [\gamma(1-p_t^i)^{\gamma-1} \log(p_t^i) + \frac{(1-p_t^i)^\gamma}{p_t^i}] * [y^i(1-p_t^i)p_t^i] \\
&= \frac{1}{n} \sum_{i=1}^n y^i(1-p_t^i)^\gamma (\gamma p_t^i \log(p_t^i) + p_t^i - 1)
\end{aligned} \tag{3.13}$$

อนุพันธ์ของ MFE

จากสมการที่ 1.6 สามารถหาอนุพันธ์ของ d เทียบกับ x ได้ดังนี้

$$\frac{\partial d}{\partial x} = d(1-d) \tag{3.14}$$

กำหนดให้

$$MSE_{major} = \frac{1}{n_{major}} \sum_{i=1}^{n_{major}} \frac{1}{2} (y^i - d^i)^2 \tag{3.15}$$

$$MSE_{minor} = \frac{1}{n_{minor}} \sum_{i=1}^{n_{minor}} \frac{1}{2} (y^i - d^i)^2 \tag{3.16}$$

ดังนั้น

$$l_{MFE} = MSE_{major} + MSE_{minor} \tag{3.17}$$

เมื่อหาอนุพันธ์ของ l_{MFE} เทียบกับ x จะได้

$$\frac{\partial l_{MFE}}{\partial x^i} = \frac{\partial MSE_{major}}{\partial x^i} + \frac{\partial MSE_{minor}}{\partial x^i} \tag{3.18}$$

โดยที่

$$\begin{aligned}
\frac{\partial MSE_{major}}{\partial x^i} &= \frac{\partial MSE_{major}}{\partial d^i} * \frac{\partial d^i}{\partial x^i} \\
&= -\frac{1}{n_{major}} \sum_{i=1}^{n_{major}} (y^i - d^i) d^i (1 - d^i)
\end{aligned} \tag{3.19}$$

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

เราจะใช้อนุพันธ์ที่ต่างกันสำหรับตัวอย่างของแต่ละกลุ่มข้อมูล กล่าวคือ ถ้าเป็นตัวอย่างของกลุ่มข้อมูลส่วนมาก สมการที่ 1.19 จะถูกใช้ และสมการที่ 1.20 จะถูกใช้เมื่อตัวอย่างของกลุ่มข้อมูลส่วนน้อย

อนุพันธ์ของ Hybrid Loss

กำหนดให้

$$FL_{major} = \frac{1}{n_{major}} \sum_{i=1}^{n_{major}} -(1 - p_t^i)^\gamma \log(p_t^i) \quad (3.21)$$

$$FL_{minor} = \frac{1}{n_{minor}} \sum_{i=1}^{n_{minor}} -(1 - p_t^i)^\gamma \log(p_t^i) \quad (3.22)$$

ดังนั้น

$$l_{Hybrid} = FL_{major} + FL_{minor} \quad (3.23)$$

เมื่อหาอนุพันธ์ของ l_{Hybrid} เทียบกับ x จะได้

$$\frac{\partial l_{Hybrid}}{\partial x^i} = \frac{\partial FL_{major}}{\partial x^i} + \frac{\partial FL_{minor}}{\partial x^i} \quad (3.24)$$

โดยที่

$$\begin{aligned} \frac{\partial FL_{major}}{\partial x^i} &= \frac{\partial FL_{major}}{\partial p_t^i} * \frac{\partial p_t^i}{\partial x^i} \\ &= \frac{1}{n_{major}} \sum_{i=1}^{n_{major}} y^i (1 - p_t^i)^\gamma (\gamma p_t^i \log(p_t^i) + p_t^i - 1) \end{aligned} \quad (3.25)$$

$$\begin{aligned} \frac{\partial FL_{minor}}{\partial x^i} &= \frac{\partial FL_{minor}}{\partial p_t^i} * \frac{\partial p_t^i}{\partial x^i} \\ &= \frac{1}{n_{minor}} \sum_{i=1}^{n_{minor}} y^i (1 - p_t^i)^\gamma (\gamma p_t^i \log(p_t^i) + p_t^i - 1) \end{aligned} \quad (3.26)$$

เช่นเดียวกับการใช้อนุพันธ์ของ MFE ก็คือ เราจะใช้สมการที่ 1.25 ถ้าตัวอย่างเป็นของกลุ่มข้อมูลส่วนมาก และจะใช้สมการที่ 1.26 ถ้าตัวอย่างเป็นของกลุ่มข้อมูลส่วนน้อย

3.2 Metrics สำหรับการประเมินประสิทธิภาพของโมเดล

3.2.1 F1-Score

F1-Score เป็นค่าเฉลี่ยของ Precision และ Recall ที่ซึ่งเราสามารถพิจารณาประสิทธิภาพของแบบจำลองด้วย F1-Score แทนการพิจารณาด้วย Precision หรือ Recall โดย F1-Score สามารถคำนวณได้จากสมการที่ 1.27

$$F1 = 2 * \left(\frac{precision * recall}{precision + recall} \right) \quad (3.27)$$

โดยที่ Precision และ Recall สามารถคำนวณได้จากสมการที่ 1.28 และ 1.29 ตามลำดับ

$$precision = \frac{TP}{TP + FP} \quad (3.28)$$

$$recall = \frac{TP}{TP + FN} \quad (3.29)$$

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

3.2.2 Receiver Operating Characteristic Curve (ROC Curve)

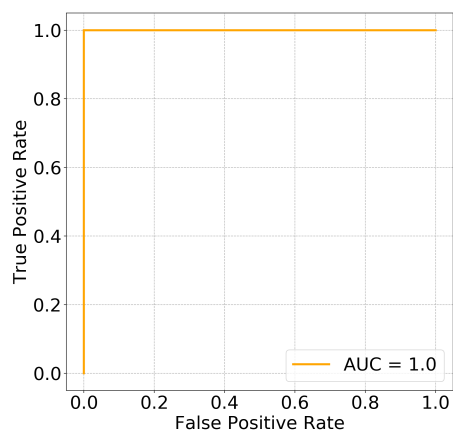
ROC Curve เป็นกราฟความสัมพันธ์ระหว่าง True Positive Rate (TPR) และ False Positive Rate (FPR) โดย TPR หรือ Recall คือ ความน่าจะเป็นที่แบบจำลองสามารถตรวจจับ Positive Class จากจำนวน Positive Class ทั้งหมด และ FPR คือ ความน่าจะเป็นที่แบบจำลองจะตรวจจับ Positive Class จากจำนวน Negative Class ทั้งหมด ทั้ง TPR และ FPR สามารถคำนวณได้จากสมการที่ 1.30 และ 1.31 ตามลำดับ

$$TPR = recall = \frac{TP}{TP + FN} \quad (3.30)$$

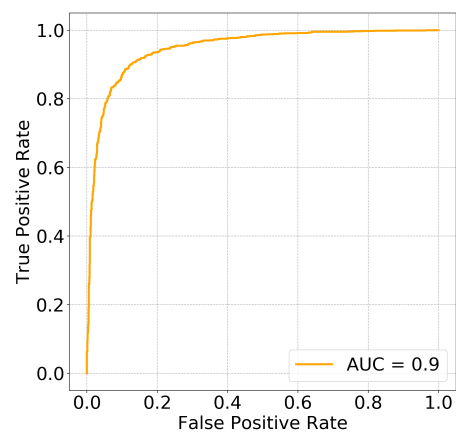
$$FPR = \frac{FP}{FP + TN} \quad (3.31)$$

3.2.3 Area Under the ROC Curve (AUC)

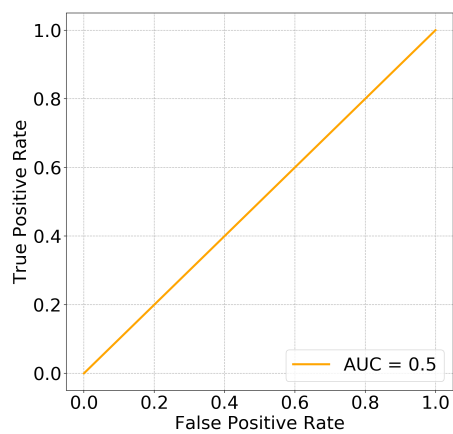
AUC คือความน่าจะเป็นที่แบบจำลองจะระบุตัวอย่างของ Positive Class ว่าเป็น Positive Class และตัวอย่างของ Negative Class ว่าเป็น Negative Class โดยถ้าค่า AUC เข้าใกล้ 1 นั้นหมายความว่าแบบจำลองมีความสามารถในการแยก Positive Class ออกจาก Negative Class ได้เป็นอย่างดี ในทางเทคนิค AUC ก็คือพื้นที่ใต้กราฟของ ROC Curve โดยความสัมพันธ์กันระหว่างค่า AUC และ ROC Curve ถูกแสดงดังรูปที่ 1.1



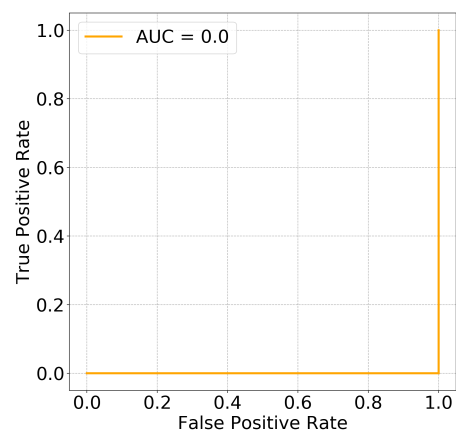
(ก)



(ข)



(ค)



(ง)

รูปที่ 3.1: ตัวอย่างกราฟ ROC Curve แบบต่าง ๆ

บทที่ 4

ผลการทดลอง

ในงานวิจัยนี้การทดลองถูกแบ่งออกเป็น 2 ส่วนหลัก ๆ คือ การจัดกลุ่มข้อมูลสองกลุ่ม (Binary Classification) และ การจัดกลุ่มข้อมูลหลายกลุ่ม (Multi-Class Classification) โดยประเภทของข้อมูลที่ใช้ในการทดลองมี 2 ประเภท คือ ข้อมูลแบบมีโครงสร้าง และ ข้อมูลภาพ สำหรับข้อมูลแบบมีโครงสร้างจะมีทั้งหมด 23 ชุดข้อมูล ที่ซึ่งแต่ละชุดข้อมูลจะมีกลุ่มของข้อมูลอยู่ 2 กลุ่ม โดยชุดข้อมูลทั้งหมดนี้จะเป็นชุดข้อมูลที่ไม่สมดุลสำหรับการวัดเปรียบเทียบสมรรถนะเกณฑ์มาตรฐาน (Benchmark) ตามที่ถูกลำดับไว้ใน [10] สำหรับข้อมูลภาพ จะมี 2 ชุดข้อมูล คือ CIFAR-100^[11] และ CIFAR-10^[11] โดยทั้งสองนี้เป็นชุดข้อมูลที่สมดุลกันอยู่แล้ว และมีจำนวนกลุ่มข้อมูล 100 และ 10 กลุ่มตามลำดับ แต่จะถูกนำมาดัดแปลงให้ไม่สมดุลเพื่อทำการทดลอง โดยรายละเอียดของการดัดแปลงของ CIFAR-100 และ CIFAR-10 จะถูกอธิบายไว้ในหัวข้อต่อไป

สำหรับสถาปัตยกรรมของโมเดลที่ใช้ในการทดลองถูกแบ่งออกเป็น 2 ส่วน คือ โมเดลโครงข่ายประสาทเทียมแบบมี 1 Hidden Layer สำหรับการทดลองกับข้อมูลแบบมีโครงสร้าง และ โมเดล ResNet^[12] สำหรับการทดลองกับข้อมูลภาพ

4.1 Binary Classification

ผลการทดลองกับชุดข้อมูลแบบมีโครงสร้าง

ผลการทดลองกับชุดข้อมูล CIFAR-100

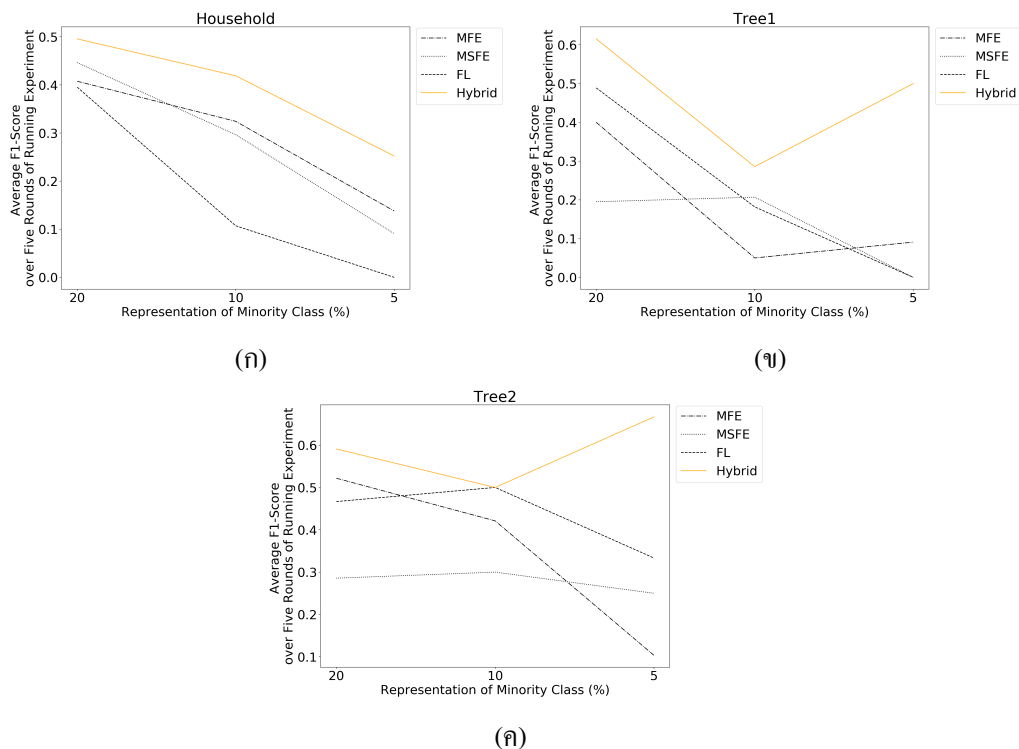
การทดลองใน [9] ได้ทำการดัดแปลง CIFAR-100 โดยการแบ่งชุดข้อมูลออกเป็น 3 ชุดข้อมูลย่อย คือ Tree1, Tree2 และ Household และเลือกข้อมูลมา 2 กลุ่มที่แตกต่างกันสำหรับแต่ละชุดข้อมูลย่อย ที่ซึ่งข้อมูลในแต่ละชุดข้อมูลย่อยจะถูกทำให้ไม่สมดุลกัน สำหรับรายละเอียดของกลุ่มข้อมูลในแต่ละชุดข้อมูลย่อยดังนี้

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

รูปที่ 2.1 แสดงผลการทดลองของแต่ละชุดข้อมูล โดยจากผลการทดลองสามารถสรุปได้ว่าฟังก์ชันสูญเสียแบบ Hybrid มีประสิทธิภาพเหนือกว่าฟังก์ชันสูญเสียอื่น ๆ อย่างสิ้นเชิง



รูปที่ 4.1: ประสิทธิภาพของแต่ละฟังก์ชันสูญเสีย เมื่อทดสอบกับชุดข้อมูลที่มีอัตราจำนวนตัวอย่างของกลุ่มข้อมูลส่วนน้อยที่แตกต่างกัน

4.2 Multi-Class Classification

ผลการทดลองกับชุดข้อมูล Long-Tailed-Imbalanced CIFAR-10

ชุดข้อมูล CIFAR-10 จะถูกดัดแปลงให้ไม่สมดุลแบบ Long-Tail

ผลการทดลองกับชุดข้อมูล Step-Imbalanced CIFAR-10

ชุดข้อมูล CIFAR-10 จะถูกดัดแปลงให้ไม่สมดุลแบบ Step

บทที่ 5

บทสรุป

บรรณานุกรม

- [1] M. Buda, A. Maki, and M. A. Mazurowski, “A systematic study of the class imbalance problem in convolutional neural networks,” *CoRR*, vol. abs/1710.05381, 2017. [Online]. Available: <http://arxiv.org/abs/1710.05381>
- [2] 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.
- [3] 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.
- [4] C. X. Ling and V. S. Sheng, *Cost-Sensitive Learning*. Boston, MA: Springer US, 2010, pp. 231–235.
- [5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [6] Y. Yan, M. Chen, M. Shyu, and S. Chen, “Deep learning for imbalanced multimedia data classification,” in *2015 IEEE International Symposium on Multimedia (ISM)*, Dec 2015, pp. 483–488.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] Z. Ding, “Diversified ensemble classifiers for highly imbalanced data learning and their application in bioinformatics,” 2011.
- [11] A. Krizhevsky, “Learning multiple layers of features from tiny images,” Tech. Rep., 2009.
- [12] 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.

ภาคผนวก ก
เรื่องที่หนึ่ง