# บทบัญญัติของกฎหมายที่ใช้บังคับในคดีละเมิด: การวิเคราะห์คำพิพากษาศาลไทยด้วยกระบวนการเรียนรู้เชิงลึก

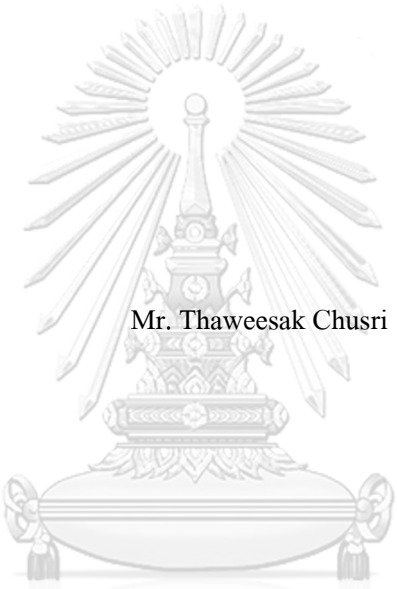
นายทวีศักดิ์ ชูศรี

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

# Applicable Provisions of Law in Tort Case: The Analysis of Thai Court Judgments using Deep Learning

Mr. Thaweesak Chusri



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Engineering in Computer Engineering

Department of Computer Engineering

FACULTY OF ENGINEERING

Chulalongkorn University

Academic Year 2022

Copyright of Chulalongkorn University

| | |
|---|---|
| หัวข้อวิทยานิพนธ์ | บทบัญญัติของกฎหมายที่ใช้บังคับในคดีละเมิด: การวิเคราะห์คำพิพากษาศาลไทยด้วยกระบวนการเรียนรู้เชิงลึก |
| โดย | นายทวีศักดิ์ ชูศรี |
| สาขาวิชา | วิศวกรรมคอมพิวเตอร์ |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก | รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม | อาจารย์ ดร.เอกพล ช่วงสุวนิช<br>รองศาสตราจารย์ ดร.อรรถพล ธำรงรัตนฤทธิ์ |

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

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

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร. โปรดปราน บุณยพุกกณะ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(อาจารย์ ดร.เอกพล ช่วงสุวนิช)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(รองศาสตราจารย์ ดร.อรรถพล ธำรงรัตนฤทธิ์)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.พีรพัฒ โชคสุวัฒนสกุล)

..... กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.ปรัชญา บุญขวัญ)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ทวีศักดิ์ ชูศรี : บทบัญญัติของกฎหมายที่ใช้บังคับในคดีละเมิด: การวิเคราะห์คำพิพากษาศาลไทยด้วยกระบวนการเรียนรู้เชิงลึก. ( Applicable Provisions of Law in Tort Case: The Analysis of Thai Court Judgments using Deep Learning ) อ.ที่ปรึกษาหลัก : รศ. ดร.อติวงศ์ สุชาโต, อ.ที่ปรึกษาร่วม : อ. ดร.เอกพล ช่วงสุวนิช,รศ. ดร.อรรถพล ธำรงรัตนฤทธิ์

การประมวลผลเอกสารทางกฎหมายโดยใช้การประมวลผลภาษาธรรมชาติเป็นงานวิจัยที่น่าสนใจในการศึกษาอย่างมาก เนื่องจากการกำหนดหมวดกฎหมายสำหรับคดีในศาลมักต้องมีการปรึกษาหารือกับทนายความซึ่งมีค่าบริการที่สูงมาก ทำให้ประชาชนทั่วไปไม่สามารถเข้าถึงการใช้บริการดังกล่าวได้ จึงเป็นจุดเริ่มต้นในการจัดทำงานวิจัยนี้ ด้วยการสร้างระบบที่สามารถดึงข้อมูลส่วนกฎหมายที่เหมาะสมตามข้อเท็จจริงที่เกี่ยวข้องของคดีในศาล

เพื่อให้บรรลุงานวิจัยนี้ ทางคณะผู้วิจัยได้รวบรวมชุดข้อมูลที่ครอบคลุมของคดีในศาลฎีกาจากประเทศไทย รวมถึงดึงข้อเท็จจริงจากเอกสารของโจทก์และจำเลยโดยใช้การผสมผสานระหว่างการเรียนรู้ของเครื่องและระบบที่ใช้กฎเป็นพื้นฐานเพื่อการวิเคราะห์การประมวลผลภาษาธรรมชาติเพิ่มเติม แนวทางที่ของงานวิจัยนี้นำเสนอมุ่งเน้นไปที่ระบบค้นคืนมาตราที่เกี่ยวข้องด้วยข้อมูลชุดฝึกที่มีจำนวนน้อย โดยใช้ข้อเท็จจริงของโจทก์เป็นข้อมูลเข้า ซึ่งระบบนี้จะสามารถจัดการกับมาตราต่างๆ ของกฎหมาย รวมถึงส่วนที่ไม่ค่อยพบหรือไม่ได้อยู่ในชุดการฝึกอบรม โดยที่ระบบจะทำงานได้ดีกว่ามาตรฐานพื้นฐาน

โดยสรุป การวิจัยนี้มีเป้าหมายเพื่อสร้างระบบดึงข้อมูลส่วนกฎหมายที่เหมาะสมตามข้อเท็จจริงที่เกี่ยวข้องของคดีในศาลที่เข้าถึงได้และแม่นยำมากขึ้น ทางคณะผู้วิจัยหวังว่าจะลดความจำเป็นในการปรึกษาทนายความที่มีค่าบริการที่สูง และมอบเครื่องมือที่มีค่าสำหรับผู้เชี่ยวชาญด้านกฎหมายและบุคคลที่เกี่ยวข้องในกระบวนการพิจารณาคดีในศาล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

| | | |
|---|---|---|
| สาขาวิชา | วิศวกรรมคอมพิวเตอร์ | ลายมือชื่อนิสิต ..... |
| ปีการศึกษา | 2565 | ลายมือชื่อ อ.ที่ปรึกษาหลัก ..... |
| | | ลายมือชื่อ อ.ที่ปรึกษาร่วม ..... |
| | | ลายมือชื่อ อ.ที่ปรึกษาร่วม ..... |

## 6170176021 : MAJOR COMPUTER ENGINEERING

KEYWORD: information retrieval, tort cases, Natural language processing

Thaweesak Chusri : Applicable Provisions of Law in Tort Case: The Analysis of Thai Court Judgments using Deep Learning . Advisor: Assoc. Prof. Dr. ATIWONG SUCHATO, Ph.D. Co-advisor: Dr. Ekapol Chuangsuwanich, Ph.D.,Assoc. Prof. Dr. ATTAPOL THAMRONGRATTANARIT, Ph.D.

Processing legal documents and understanding arrays of law sections has proven to be a challenging task for natural language processing. Determining the relevant law section for a court case often requires expensive consultations with lawyers, which many people cannot afford. Our goal is to create a system that can retrieve the appropriate law sections based on the relevant facts of a court case.

To achieve this, we have gathered a comprehensive dataset of Supreme Court Cases from Thailand. Using a combination of machine learning and rule-based systems, we extract the facts from the plaintiff's and defendant's documents, enhancing the document's structure for further NLP analysis. Our proposed approach focuses on a few-shot law retrieval system, using plaintiff facts as input. This allows us to handle the wide range of law sections, including those that are rarely encountered or not present in the training set. Our system performs better than the standard supervised baseline and can handle previously unseen law sections.

In summary, our research aims to address the difficulties in processing legal documents and determining relevant law sections for court cases. By creating a more accessible and accurate system, we hope to reduce the need for expensive lawyer consultations and provide a valuable tool for legal professionals and individuals involved in court proceedings.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

| | | |
|---|---|---|
| Field of Study: | Computer Engineering | Student's Signature ..... |
| Academic Year: | 2022 | Advisor's Signature ..... |
| | | Co-advisor's Signature ..... |
| | | Co-advisor's Signature ..... |

# กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้คงไม่อาจสำเร็จลุล่วงไปได้หากไม่ได้รับความอนุเคราะห์จากรองศาสตราจารย์ ดร.อติวงศ์ สุชาโต ที่กรุณาเสียสละเวลาอันมีค่ารับเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งได้คอยช่วยเหลือ ให้คำปรึกษา และให้ความรู้ที่มีคุณค่าในการเขียนวิทยานิพนธ์นี้จนสำเร็จและผ่านพ้นไปได้ด้วยดี

ขอขอบพระคุณ อาจารย์ ดร.เอกพล ช่วงสุวนิช และ รศ.ดร.อรรถพล ธำรงรัตนฤทธิ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ซึ่งเป็นผู้ที่คอยให้คำแนะนำและให้ความรู้ทั้งในแง่การให้ความรู้ทางการวิจัย องค์ความรู้ด้านการประมวลผลภาษาธรรมชาติร่วมกับเทคนิคการเรียนรู้เชิงลึก ซึ่งมีส่วนสำคัญต่อความสำเร็จของวิทยานิพนธ์ฉบับนี้ในการจัดทำวิทยานิพนธ์รวมถึงการตีพิมพ์งานวิจัยชิ้นนี้ด้วย

ขอขอบพระคุณ รศ.ดร.โปรดปราน บุณยพุกกะ ที่กรุณาเสียสละเวลาอันมีค่ารับเป็นประธานกรรมการสอบวิทยานิพนธ์ ท่านอาจารย์ได้กรุณาให้คำแนะนำที่มีค่าอย่างยิ่งต่อผู้วิจัย

ขอขอบพระคุณ ผศ.ดร.พีรพัฒ โชตสุวัตนสกุล กรรมการสอบวิทยานิพนธ์ที่ได้มอบองค์ความรู้อันมีค่าทางด้านกฎหมาย เพื่อให้องค์ความรู้ทางวิทยาศาสตร์คอมพิวเตอร์และกฎหมายสามารถประสานรวมกันกลายเป็นวิทยานิพนธ์ฉบับนี้ได้สำเร็จ

ขอขอบพระคุณ อ.ดร.ปรัชญา บุญขวัญ กรรมการสอบวิทยานิพนธ์ ที่ได้กรุณาให้คำแนะนำและความรู้ที่มีค่าอย่างยิ่งให้แก่ผู้วิจัยในการเขียนวิทยานิพนธ์นี้จนสำเร็จ

รวมถึงขอขอบคุณภาควิชาคณะวิศวกรรมคอมพิวเตอร์ และเจ้าหน้าที่จุฬาลงกรณ์มหาวิทยาลัยทุกท่านที่ให้ความช่วยเหลือ และสนับสนุนจนวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี

สุดท้ายนี้ข้าพเจ้าขอขอบพระคุณผู้ที่เกี่ยวข้องท่านอื่นๆที่ไม่ได้กล่าวมาในข้างต้น ซึ่งมีผลให้วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดี ข้าพเจ้าหวังเป็นอย่างยิ่งว่า วิทยานิพนธ์ฉบับนี้จะสามารถสร้างคุณประโยชน์ให้แก่ผู้ที่สนใจศึกษาได้ไม่มากก็น้อย และหากวิทยานิพนธ์ฉบับนี้มีข้อผิดพลาดประการใดต้องขออภัยมา ณ ทีนี้ ด้วย

ทวีศักดิ์ ชูศรี

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

# สารบัญ

หน้า


บทคัดย่อภาษาไทย ..... ค

บทคัดย่อภาษาอังกฤษ ..... ง

กิตติกรรมประกาศ ..... จ

สารบัญ ..... ฉ

สารบัญตาราง ..... ฌ

สารบัญรูปภาพ ..... ญ

บทที่ 1 ..... 1

บทนำ ..... 1

1.1 ที่มาและความสำคัญ ..... 1

1.2 วัตถุประสงค์ ..... 2

1.3 ขอบเขตการดำเนินงาน ..... 2

1.4 ขั้นตอนการดำเนินงาน ..... 3

1.5 ประโยชน์ที่คาดว่าจะได้รับ ..... 3

1.6 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์ ..... 3

1.7 ผลงานที่ได้รับการตีพิมพ์จากวิทยานิพนธ์ ..... 4

บทที่ 2 ..... 5

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง ..... 5

2.1 ทฤษฎีที่เกี่ยวข้อง ..... 5

2.1.1 การฝังคำ (Word Embedding) ..... 5

การฝังคำ (Word Embedding) ..... 5

2.1.1.1 วิธีการนับจำนวน (Count-based Methods) ..... 5



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.1.1.2 วิธีการเชิงทำนาย (Predictive Methods).....6
2.1.2 การถดถอยโลจิสติก (Logistic Regression).....6
2.1.2.1 ช่วงการปรับค่าน้ำหนัก.....6
2.1.2.2 ช่วงการทดสอบแบบจำลอง.....6
2.1.3 Deep Averaging network (DAN).....8
2.1.3.1 Word Embedding.....9
2.1.3.2 Averaging.....9
2.1.3.3 Deep Layers.....9
2.1.3.4 Classification.....9
2.1.4 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network).....9
2.1.5 โครงข่ายประสาทเทียมแบบวงวน (Recurrent Neural Network).....11
2.1.6 แบบจำลองตัวเข้ารหัสแบบสองทิศทางในรูปแบบตัวแปลง (Bidirectional Encoder Representations from Transformers: BERT).....14
2.1.7 การเรียนรู้ด้วยข้อมูลจำนวนน้อย (Few-Shot Learning) และการเรียนรู้โดยปราศจากข้อมูลปฐมภูมิ (Zero-Shot Learning).....20
2.1.8 การประเมินผลแบบจำลอง.....22
2.1.8.1 Accuracy (ความแม่นยำ).....22
2.1.8.2 Precision (ความเที่ยงตรง).....23
2.1.8.3 Recall (ความถูกต้อง).....23
2.1.8.4 F1-score (คะแนน F1).....23
2.2 งานวิจัยที่เกี่ยวข้อง.....24
2.2.1 งานวิจัยเกี่ยวกับการวิเคราะห์คำพิพากษาของศาลในต่างประเทศ.....24
2.2.2 งานวิจัยเกี่ยวกับคำพิพากษาของศาลในประเทศไทย.....26
บทที่ 3 แนวคิดและวิธีการวิจัย.....28

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.1 การสร้างคลังข้อมูลคำพิพากษา....28
3.2 ขั้นตอนการรวบรวมคำพิพากษา....28
3.3 การคัดแยกคำพิพากษาที่เกี่ยวข้องกับคดีละเมิด....29
3.3.1 ข้อมูลนำเข้าในส่วนของคำฟ้อง....31
3.3.2 ข้อมูลนำเข้าในส่วนของคำให้การ....33
3.3.3 ข้อมูลนำเข้าในส่วนตัวบทบัญญัติของกฎหมาย....34
3.3 รายละเอียดการทดลอง....36
3.3.1 การจัดเตรียมชุดข้อมูล....36
3.3.2 รูปแบบสถาปัตยกรรมในการทดลอง....39
3.3.2.1 ใช้คำฟ้องเท่านั้นเป็นค่าสัญญาณ....40
3.3.2.2 ใช้คำฟ้องและคำให้การเป็นค่าสัญญาณ....40
3.3.2.3 สถาปัตยกรรมโครงข่ายประสาทเทียมสยาม (Siam Neural Network)....41
3.3.3 การออกแบบการทดลอง....44
3.3.3.1 การแบ่งข้อมูล....44
3.3.3.2 การปรับจูนไฮเปอร์พารามิเตอร์....45
บทที่ 4....48
ผลการวิจัย....48
บทที่ 5....52
สรุปผล....52
บรรณานุกรม....53
ประวัติผู้เขียน....57

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

# สารบัญตาราง

**หน้า**

ตาราง 1 แสดงโครงสร้างของการจัดเก็บข้อมูลคำพากษา.....28

ตาราง 2 โครงสร้างการเก็บข้อมูลตัวบทบัญญัติของกฎหมาย .....35

ตาราง 3 จำนวนโทเคนในข้อมูลรับเข้าแต่ละประเภท .....38

ตาราง 4 ตารางแสดงตัวอย่างรูปแบบข้อมูลของแบบจำลองที่ใช้เฉพาะคำฟ้องเป็นค่าสัญญาณ .....40

ตาราง 5 ตารางแสดงตัวอย่างรูปแบบข้อมูลของแบบจำลองที่ใช้ทั้งคำฟ้องและคำให้การเป็นค่าสัญญาณ.....40

ตาราง 6 ตารางแสดงตัวอย่างรูปแบบข้อมูลของแบบจำลองในโครงข่ายย่อยแบบ Few-shot .....42

ตาราง 7 ตารางแสดงตัวอย่างรูปแบบข้อมูลของแบบจำลองในโครงข่ายย่อยแบบ Zero-shot .....43

ตาราง 8 แสดงผลการทดลองในรูปแบบของ f1-score ของการจำแนกข้อมูลสองประเภทโดยเปรียบเทียบระหว่างแบบจำลองและระหว่างลักษณะข้อมูลรับเข้า โดยกำหนดให้ P หมายถึงคำฟ้องและ D หมายถึงคำให้การ .....48

ตาราง 9 ผลการเปรียบเทียบค่า f1-score ของแต่ละสถาปัตยกรรม .....49

ตาราง 10 ผล f1-score@3 ของโครงข่ายประสาทเทียมสยามโดยเปรียบเทียบผลระหว่างแบบจำลอง .....51

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

# สารบัญรูปภาพ

หน้า

รูปภาพ 1 Deep Averaging Network ที่มา (Iyyer, Manjunatha et al. 2015) .....8

รูปภาพ 2 กระบวนการใช้แบบจำลอง CNN กับข้อความ (Zhang and Wallace 2015).....10

รูปภาพ 3 แผนภาพการทำงานของแบบจำลอง Vanilla RNN.....12

รูปภาพ 4 แผนภาพแบบจำลอง LSTM (Ismail, Wood et al. 2018) .....13

รูปภาพ 5 LSTM มีการประมวลผลคำแต่ละคำ โดยใช้สมการดังนี้ (Ismail, Wood et al. 2018) .....13

รูปภาพ 6 แผนภาพการคำนวณเวกเตอร์ Query เวกเตอร์ Key และเวกเตอร์ Value ของแต่ละคำที่ปรากฏในข้อความ (Alammar 2018) .....15

รูปภาพ 7 แผนภาพการคำนวณ attention ของคำที่อยู่ใกล้เคียง (Alammar 2018).....15

รูปภาพ 8 แผนภาพการประกอบสร้าง Word Embedding ใหม่จาก Attention และเวกเตอร์ Value (Alammar 2018).....16

รูปภาพ 9 แผนภาพการเข้ารหัสของตำแหน่งคำ และการบวกและการปัดให้เป็นค่าปกติ (Alammar 2018) .....17

รูปภาพ 10 แผนภาพกระบวนการ Pretraining แบบจำลอง BERT โดยการใช้ Masked Language Modeling Task (Alammar 2018).....18

รูปภาพ 11 แผนภาพกระบวนการ Fine-tuning เพื่อถ่ายทอดความรู้จากแบบจำลองที่ถูกเทรนมาก่อนส่วนหนึ่ง (Alammar 2018).....19

รูปภาพ 12 แผนภาพแสดงการทำงานของแบบจำลอง (Koch, Zemel et al. 2015).....21

รูปภาพ 13 สถิติมาตรากฎหมายที่เกี่ยวข้องกับคดีละเมิดจากวิเคราะห์คลังข้อมูล .....30

รูปภาพ 14 แสดงตัวอย่างของข้อมูลนำเข้าจำนวนหนึ่งเอกสาร .....31

รูปภาพ 15 แสดงโครงสร้างของแบบจำลองในการระบุตำแหน่งเริ่มต้นของคำให้การ .....32

รูปภาพ 16 แสดง Regular Expression ของการแยกคำให้การ .....33

รูปภาพ 17 แสดง Regular Expression ที่สามารถค้นหาและแยกข้อความได้.....34

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

รูปภาพ 18 หน้าจอแสดงผลการค้นหาประมวลกฎหมายจากเว็บไซต์สำนักงานคณะกรรมการกฤษฎีกา .....34
รูปภาพ 19 ลิงก์ของ ประมวลกฎหมายแพ่งและพาณิชย์ (ฉบับ Update ล่าสุด).....35
รูปภาพ 20 หน้าหลักของประมวลกฎหมายแพ่งและพาณิชย์ (ฉบับ Update ล่าสุด) .....35
รูปภาพ 21 แผนภาพแบบจำลองแบบจำแนกข้อมูลสองประเภท .....39
รูปภาพ 22 แผนภาพแบบจำลองแบบโครงข่ายประสาทเทียมสยาม .....41
รูปภาพ 23 สัดส่วนป้ายกำกับในแบบจำลองการจำแนกแบบสองประเภทของแต่ละมาตรา .....44
รูปภาพ 24 สัดส่วนป้ายกำกับในแบบจำลองแบบโครงข่ายประสาทเทียมสยามของแต่ละมาตรา.45
รูปภาพ 25 แผนภาพแบบจำลอง mBERT แบบจำแนกข้อมูลสองประเภท.....46
รูปภาพ 26 แผนภาพแบบจำลอง mBERT แบบโครงข่ายประสาทเทียมสยาม.....47

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

# บทที่ 1
## บทนำ

### 1.1 ที่มาและความสำคัญ

ในปัจจุบันการเข้าถึงทนายความที่มีคุณภาพของประชาชนทั่วไปมีข้อจำกัดอย่างมาก เนื่องจากจำนวนทนายความที่มีอยู่อาจไม่เพียงพอต่อความ ต้องการอีกทั้งยังมีค่าบริการที่สูงซึ่งทำให้มีประชาชนทั่วไปจำนวนมากที่ ไม่สามารถ เข้าถึงข้อมูลที่ จำเป็น เกี่ยวกับสิทธิ หน้าที่และความรับผิดชอบ ทางกฎหมาย ที่ตนมีอยู่รวมถึงการแก้ไขข้อพิพาททางกฎหมายที่อาจเกิดขึ้น หรือ ที่กำลังเกิดขึ้นอยู่

ซึ่งเป็นสาเหตุสำคัญที่ทำให้เกิดคดีตกค้างในศาลมีจำนวนมากเป็นเหตุมาจากกระบวนการพิจารณาคดีที่ใช้เวลานานและจำนวนข้อเท็จจริงของ คดีที่ศาลต้องพิจารณามีมากเกินกว่าที่ศาลสามารถตัดสินได้ในระยะเวลาที่กำหนดไว้ เนื่องจากสถานการณ์ดังกล่าวจึงจำเป็นต้องมีวิธีที่ช่วยให้คดีเหล่านี้ สามารถยุติหรือแก้ไข ระหว่างคู่ความได้โดยไม่จำเป็นต้องดำเนินเรื่องไปสู่ ชั้นศาล

จากการศึกษาพบว่ามีเพียงบางส่วนของข้อพิพาทที่จำเป็นต้องนำสู่การพิจารณาของศาลซึ่งปัจจัยสำคัญที่ช่วยในการกำหนดว่า ข้อพิพาทหนึ่ง ๆ จะถูกยุติโดยศาลหรือสามารถแก้ไขระหว่างคู่ความได้คือ ความคาดหมาย เกี่ยวกับผลลัพธ์ที่จะได้รับจากคำตัดสินของศาลในข้อพิพาทนั้น กล่าวคือ หากคู่ความมีความเห็นที่ตรงกันเกี่ยวกับผลคำพิพากษาที่อาจเกิดขึ้นหากข้อพิพาทนั้นถูกนำสู่ศาล ทำให้สามารถตกลงกันได้เพื่อประหยัดค่าใช้จ่าย รวมถึงลดต้นทุนและค่าเสียโอกาสอื่น ๆ ที่เกิดจากการนำคดีไปสู่ศาล ในกรณี ที่เป็นเช่นนั้นข้อพิพาทที่มีลักษณะคล้ายคลึงกับคดีที่ศาลเคยมีคำพิพากษา ในทิศทางของเรื่องนั้น ๆ อาจไม่จำเป็นต้องส่งขึ้นกระบวนพิจารณาคดี อีกครั้งซึ่งจะช่วยให้ศาลสามารถใช้เวลาและทรัพยากรที่มีอยู่ในการพิจารณาคดี ที่ซับซ้อนและสำคัญมากกว่าได้ นอกจากนี้ยังต้องใช้ความรู้ความสามารถ และความเชี่ยวชาญของศาลในการตัดสินคดีอย่างถูกต้องและยุติสิ้นเชิง

จากการสำรวจสถิติคดีที่เข้าสู่กระบวนการพิจารณาในศาลในหลายปีที่ผ่านมา พบว่าคดีละเมิด (Tort)เป็นหนึ่งในประเภทของคดีที่มีจำนวนมาก ที่สุดที่เข้าสู่ชั้นศาลดังนั้น คดีเหล่านี้เป็นจุดเริ่มต้นที่ดีที่จะนำเสนอเพื่อ วิเคราะห์ ด้วย การประมวลผลภูมิปัญญาประดิษฐ์เพื่อสร้างแบบจำลอง (Model) ที่สามารถทำนายผลของคำพิพากษาของศาลในคดีละเมิดได้ โดยในขั้นต้นคณะผู้วิจัยมีเป้าหมายที่จะพัฒนาแบบจำลองเพื่อทำนาย ข้อกฎหมายที่เกี่ยวข้องกับคำพิพากษา เพื่อเป็นฐานในการพัฒนาแบบจำลองทำนายผล ของคำพิพากษาในอนาคต

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

**1.2 วัตถุประสงค์**

1. เพื่อศึกษาและสร้างคลังข้อมูลคำพิพากษาศาลฎีกาเพื่อให้นักวิจัยสามารถศึกษาและนำไปใช้ประโยชน์ในการวิจัยต่อไป
2. เพื่อศึกษาและสร้างพัฒนาแบบจำลองที่สามารถใช้ในการเลือกกฎหมายที่เกี่ยวข้องในคดีละเมิดจากคำให้การ

**1.3 ขอบเขตการดำเนินงาน**

1. ข้อมูลที่ใช้ในการสร้างคลังข้อมูลคำพิพากษาศาลฎีกานั้นมา จากการรวบรวมกจากข้อมูลที่ถูกเผยแพร่ในเว็บไซต์ระบบสืบค้นคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา(ศูนย์เทคโนโลยีสารสนเทศและการสื่อสารในศาลฎีกา 2017) ทั้งนี้ในการทดลองของแบบจำลองจะใช้เฉพาะคำพิพากษาที่เกี่ยวข้องกับคดีละเมิดหรือก็คือคดีที่มีความเกี่ยวข้องกับประมวลกฎหมายแพ่งและพาณิชย์ว่าด้วยละเมิดมาตรา 420
2. การพัฒนาแบบจำลองเพื่อหามาตราที่เกี่ยวข้อง ส่วนข้อมูลนำเข้า (input) จะประกอบด้วย 3 ประเภทตามที่ได้กล่าวไว้ดังนี้:

2.1. คำฟ้อง (Plaint): เป็นส่วนของคำพิพากษาที่ถูกยกขึ้นโดยผู้กล่าวหา ซึ่งประกอบด้วยข้อกล่าวหาและข้อมูลที่เกี่ยวข้อง เช่น ความผิดที่ถูกกล่าวหาและบทลงโทษที่อาจจะประกอบคำพิพากษา

2.2. คำให้การ (Defense): เป็นส่วนของคำพิพากษาที่ถูกยกมาโต้แย้งโดยผู้ถูกกล่าวหา ซึ่งประกอบด้วยข้อโต้แย้งและข้อมูลที่เกี่ยวข้อง เช่น การปรากฏฐานหรือเหตุที่สอดคล้องกับคำฟ้อง

2.3. ตัวบทกฎหมาย (Legal Statute): เป็นส่วนที่ถูกบัญญัติขึ้นและระบุมาตราที่เกี่ยวข้องกับคดี ซึ่งสามารถใช้เพื่อทำความเข้าใจเกี่ยวกับบทลงโทษที่อาจจะมีผลในการพิจารณาคดี

3. ในการทดลองสร้างแบบจำลองนั้นจะทำการทดลองด้วยกัน 4 แบบจำลอง

3.1. Binary Classification (One vs All): แบบจำลองนี้ใช้ในการจำแนกข้อมูลออกเป็นสองกลุ่ม โดยใช้เทคนิค One vs All ซึ่งแยกข้อมูลในกลุ่มที่สนใจกับกลุ่มอื่นๆ

3.2. การเรียนรู้ด้วยข้อมูลจำนวนน้อย (Few-Shot Learning) โดยใช้แบบจำลอง Siamese Network: แบบจำลอง Siamese Network ใช้ในการเรียนรู้จากจำนวนข้อมูลที่จำกัด (Few Shot) โดยทำการเปรียบเทียบความคล้ายคลึงของคู่ของข้อมูล และตัดสินใจว่าสองข้อมูลนั้นคล้ายคลึงกันหรือไม่ ในที่นี้ใช้ในการหามาตราที่เกี่ยวข้อง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.3. การเรียนรู้โดยปราศจากข้อมูลปฐมภูมิ (Zero-Shot Learning) โดยใช้แบบจำลอง Siamese Network: แบบจำลอง Siamese Network ใช้ในการทำนายข้อมูลที่ไม่เคยเห็นมาก่อน (Zero Shot) โดยใช้การเรียนรู้จากข้อมูลที่มีอยู่เพื่อทำนายสิ่งที่ไม่เคยเห็นมาก่อน ในที่นี้ใช้ในการหามาตราที่เกี่ยวข้อง

3.4. การโอนย้ายการเรียนรู้ (Transfer Learning) โดยใช้ BERT (Bidirectional Encoder Representations from Transformers) ซึ่งเป็นแบบจำลองภาษาธรรมชาติที่ถูกฝึกสอนด้วยข้อมูลใหญ่ และสามารถนำมาใช้ในงานต่างๆ โดยแบบจำลอง BERT สามารถทำนายส่วนหนึ่งของข้อความเพื่อหาคำตอบหรือความเหมาะสม การใช้ Transfer Learning จะช่วยให้สามารถนำความรู้จากแบบจำลองที่ฝึกสอนแล้วมาใช้ในงานที่มีข้อมูลน้อย และให้ประสิทธิภาพสูงขึ้น

**1.4 ขั้นตอนการดำเนินงาน**

1. รวบรวมข้อมูลสำหรับทำวิจัย
2. ศึกษาทฤษฎีที่เกี่ยวข้อง
3. ศึกษาคนคว้างานวิจัยที่เกี่ยวข้อง
4. ออกแบบแนวคิดสำหรับการวิจัย
5. จัดเตรียมคำพิพากษาในรูปของข้อมูลนำเข้า
6. ออกแบบและทำการทดลองแบบจำลองพื้นฐาน
7. ทดสอบและประเมินผลการทดลองแบบจำลองพื้นฐาน
8. ออกแบบและนำเสนอแบบจำลองใหม่
9. ทดสอบและประเมินผลการทดลองแบบจำลองใหม่
10. สรุปผลการทดลอง
11. เผยแพร่งานวิจัย
12. จัดทำเอกสารวิทยานิพนธ์

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

**1.5 ประโยชน์ที่คาดว่าจะได้รับ**

เพื่อพัฒนาแบบจำลองเพื่อทำนายข้อกฎหมายที่เกี่ยวข้องกับคำพิพากษาในคดีละเมิดโดยการระบุข้อเท็จจริง พร้อมทั้งเผยแพร่ชุดข้อมูลคำพิพากษาที่พร้อมสำหรับการนำไปวิจัยหรือสร้างระบบอื่นๆ ต่อไปในอนาคต

**1.6 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์**

เนื้อหาในวิทยานิพนธ์แบ่งออกเป็น 5 บท ได้แก่ บทที่ 1 บทนำ อธิบายถึงที่มาและความสำคัญของปัญหา วัตถุประสงค์ของงานวิจัย ขอบเขตงานวิจัย ประโยชน์ที่คาดว่าจะได้รับ และ

ผลงานที่ได้รับการตีพิมพ์จากวิทยานิพนธ์ บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง บทที่ 3 แนวคิดและวิธีการวิจัย บทที่ 4 การออกแบบและพัฒนาเครื่องมือ และบทที่ 5 การทดสอบและการวิเคราะห์ผล

**1.7 ผลงานที่ได้รับการตีพิมพ์จากวิทยานิพนธ์**

ส่วนหนึ่งของวิทยานิพนธ์ฉบับนี้ได้รับการตีพิมพ์บทความทางวิชาการ

Thaweesak Chusri, Sadanan Arsaibun, Peerapat Chokesuwattanaskul, Ekapol Chuangsuwanich, and Attapol T. Rutherford (2023). Few-Shot Law Retrieval System for Supreme Court Cases. Proceedings of The 20th International Joint Conference on Computer Science and Software Engineering (JCSSE2023), (page 84-89).

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

# บทที่ 2
# ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

## 2.1 ทฤษฎีที่เกี่ยวข้อง

### 2.1.1 การฝังคำ (Word Embedding)

การฝังคำ (Word Embedding) เป็นกระบวนการแปลงคำศัพท์หรือประโยคในภาษาธรรมชาติเป็นเวกเตอร์ตัวเลขในเชิงคณิตศาสตร์ เพื่อให้สามารถนำคำหรือประโยคที่มีความหมายคล้ายคลึงกันในมุมมองคณิตศาสตร์ โดยมีจุดประสงค์ในการรักษาความหมายและความสัมพันธ์ของคำ หรือประโยคในการประมวลผลภาษาธรรมชาติหรืองานที่เกี่ยวข้องกับภาษาธรรมชาติ เช่น การจำแนกหมวดหมู่ของเอกสาร (Document Classification) การแปลภาษา (Machine Translation) หรือการสร้างแบบจำลองภาษา ธรรมชาติ อื่น ๆ

ในกระบวนการฝังคำจะสร้างเวกเตอร์ที่แทนคำโดยให้คำที่มีความหมายคล้ายคลึงกันมีเวกเตอร์ที่อยู่ใกล้กันและคำที่ไม่เกี่ยวข้องกันจะมีเวกเตอร์ที่อยู่ห่างกัน ด้วยเวกเตอร์ที่แทนชุดคำศัพท์เหล่านี้ จะสามารถนำมาวิเคราะห์ ความคล้ายคลึงระหว่างคำหรือประโยค หรือนำมาใช้ในการเรียนรู้และสร้างแบบจำลองที่มีประสิทธิภาพในงานที่เกี่ยวข้องกับภาษาธรรมชาติ เช่น การจำแนกหมวดหมู่ของข้อมูล (Document Classification) การแปลภาษา (Machine Translation) หรือการสร้างแบบจำลองภาษาธรรมชาติอื่น ๆ

การฝังคำสามารถแบ่งออกเป็น 2 ประเภทหลัก คือวิธีการนับจำนวน (Count-based Methods) และวิธีการเชิงทำนาย (Predictive Methods)

*2.1.1.1 วิธีการนับจำนวน (Count-based Methods)*

ใช้ในการสร้าง Word Embedding โดยใช้เทคนิคต่าง ๆ เช่น การสร้างดัชนีแฝงหรือการวิเคราะห์ความหมายแฝง (Latent Semantic Indexing/Analysis: LSI/LSA) โดยความหมายของ คำหนึ่งคำจะถูกกำหนดโดยคำที่เกิดรอบ ๆ ซึ่งเก็บข้อมูลความถี่ของคำที่เกิดร่วมไว้ในเวกเตอร์ ข้อดีของวิธีนี้คือสามารถให้คำที่ คล้ายคลึงกันมีเวกเตอร์ที่อยู่ใกล้กัน แต่มีขนาดใหญ่และ มีค่าส่วนใหญ่จะเป็นศูนย์ทำให้ใช้งานลำบากได้ในบางกรณี

*2.1.1.2 วิธีการเชิงทำนาย (Predictive Methods)*

จะใช้แบบจำลองเรียนรู้คำจากการทำนายคำที่อยู่รอบ ๆ เพื่อให้ได้เวกเตอร์ที่มีขนาดเล็กและไม่มีค่าส่วนใหญ่เป็นศูนย์ (Dense Vector) สามารถใช้แบบจำลองที่ใช้เทคนิคของโครงข่าย ประสาทเทียมได้ เช่น Word2Vec, GloVe และ FastText เวกเตอร์ที่ได้จากวิธีการเชิงทำนายมีคุณสมบัติที่ดีในการแสดงความหมายของคำและประโยค และมักนำไปใช้ในงานประมวล ผลภาษาธรรมชาติต่าง ๆ

**2.1.2 การถดถอยโลจิสติก (Logistic Regression)**

การถดถอยโลจิสติก (Logistic Regression) หรือเอนโทรปีสูงสุด (Maximum Entropy: MaxEnt) เป็นแบบจำลองการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Machine Learning) ที่ได้รับความนิยมในการสร้างเครื่องจำแนกประเภทข้อความ (Text Classifier) เนื่องจากสามารถแสดงความสำคัญของค่าสัญญาณแต่ละค่าที่ แบบจำลองเรียนรู้ได้อย่างชัดเจน และง่ายต่อการปรับแต่งพารามิเตอร์ของแบบจำลองในงานวิจัย (Jurafsky and Martin 2008) การฝึกแบบจำลองการถดถอยโลจิสติก สามารถแบ่งเป็น 2 ช่วงได้ดังนี้:

*2.1.2.1 ช่วงการปรับค่าน้ำหนัก*

ในขั้นตอนนี้จะกำหนดค่าน้ำหนัก (weights) และค่าเอนเอียง (bias term) ของแบบจำลองเพื่อให้แบบจำลองสามารถปรับค่าสัญญาณที่ได้จากข้อมูลเข้าใกล้ค่าที่เหมาะสมที่สุดกับผลลัพธ์ที่ต้องการ

*2.1.2.2 ช่วงการทดสอบแบบจำลอง*

หลังจากได้ค่าน้ำหนักและเอนเอียงที่เหมาะสมแล้ว แบบจำลองจะถูกทดสอบด้วยชุดข้อมูลทดสอบเพื่อหาความน่าจะเป็นที่ตัวอย่างข้อความจะอยู่ในกลุ่มที่กำหนดไว้ ผลลัพธ์ที่ได้จากแบบจำลองคือความน่าจะเป็นที่ข้อความจะอยู่ในกลุ่มที่กำหนด โดยใช้ฟังก์ชันซิกมอยด์ (Sigmoid function) ในการคำนวณ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

แบบจำลองการถดถอยโลจิสติก ใช้ฟังก์ชันซิกมอยด์ เพื่อหาความน่าจะเป็นผลลัพธ์ที่ได้จากแบบจำลองจะเป็นไปได้เพียง 2 ค่า (Binary) ซึ่งสามารถแทนค่านี้ด้วยตัวแปร y ซึ่งจะมีค่าเป็น 1 หากข้อมูลมีความน่าจะเป็นที่จะอยู่ในกลุ่มที่กำหนด และค่าเป็น 0 หากข้อมูลไม่มีความน่าจะเป็นที่จะอยู่ในกลุ่มที่กำหนด แบบจำลอง Logistic Regression สามารถทำนายค่าเหล่านี้ได้จากค่าน้ำหนัก (Weight: w) และค่าความเอนเอียง (Bias Term: b) ที่ถูกปรับในระหว่างกระบวนการฝึกแบบจำลอง โดยค่าน้ำหนักในที่นี้หมายถึงความสำคัญของค่าสัญญาณของข้อมูลรับเข้าที่แบบจำลอง กำลังเรียนรู้อยู่ และเป็นค่าที่บ่งบอกให้แบบจำลอง รู้ว่าข้อมูลนั้นควรเป็นค่าสัญญาณที่ชี้ว่าเป็นผลเชิงบวก (Positive Label)

หรือค่าสัญญาณที่ชี้ว่าเป็นผลเชิงลบ (Negative Label) ส่วนค่าความเอนเอียงเป็นค่าจำนวนจริงที่บวกเข้าไปในค่าของข้อมูลรับเข้าที่ถ่วงน้ำหนักแล้วในตอนสุดท้าย ค่าน้ำหนักและค่าความเอนเอียงจะถูกปรับให้เหมาะสมในระหว่างกระบวนการฝึกแบบจำลอง ดังนั้น สมการที่แสดงการทำงานของแบบจำลองการถดถอยโลจิสติก มีดังนี้

$$z = (\sum_{i=1}^{n} w_i x_i) + b \qquad (1)$$

ให้ [n] แทนจำนวนข้อมูลทั้งหมดที่ถูกป้อนเข้ามาให้แบบจำลอง
และ (i) แทนลำดับปัจจุบันของข้อมูล

หลังจากแบบจำลองได้ฝึกและปรับค่าน้ำหนัก (w) แล้ว แบบจำลอง จะใช้ค่าน้ำหนักนี้คูณกับแต่ละค่าของข้อมูลรับเข้า (x) เพื่อคำนวณค่าที่ถ่วงน้ำหนักแล้วของสัญญาณที่มี จากนั้นแบบจำลองจะหาผลรวมของค่าที่ถ่วงน้ำหนักแล้วทั้งหมดของสัญญาณนี้ แล้วนำค่านี้ไปบวกกับค่าความเอนเอียง (b) โดยที่ z เป็นคะแนนที่ป้ายกำกับนั้นได้รับ ส่วนสำคัญอย่างยิ่งในการฝึกแบบจำลองการถดถอยโลจิสติก คือการใช้ฟังก์ชันสูญเสีย (Loss Function) หรือฟังก์ชันต้นทุน (Cost Function) เพื่อวัดความต่างระหว่างความน่าจะเป็นที่แบบจำลองทำนายกับความเป็นจริงของป้ายกำกับฟังก์ชันสูญเสียที่นิยมใช้ใน Logistic Regression และแบบจำลอง ประสาทเทียมคือฟังก์ชันครอส-เอนโทรปี (Cross-entropy Loss) เมื่อแบบจำลองคำนวณความน่าจะเป็นของแต่ละป้ายกำกับในข้อมูลและนำผลลัพธ์ไปแปลงเป็นค่าลอการิทึม (Logarithm) จะได้ค่าที่เรียกว่าลอการิทึมของความเป็นไปได้ (Log-likelihood) จากสมการดังต่อไปนี้:

Chulalongkorn University

Log-likelihood

$$\arg\max_W \ell(W; X, Y) = \arg\min_W -\ell(W; X, Y)$$

Crossentropy loss
Negative log-likelihood

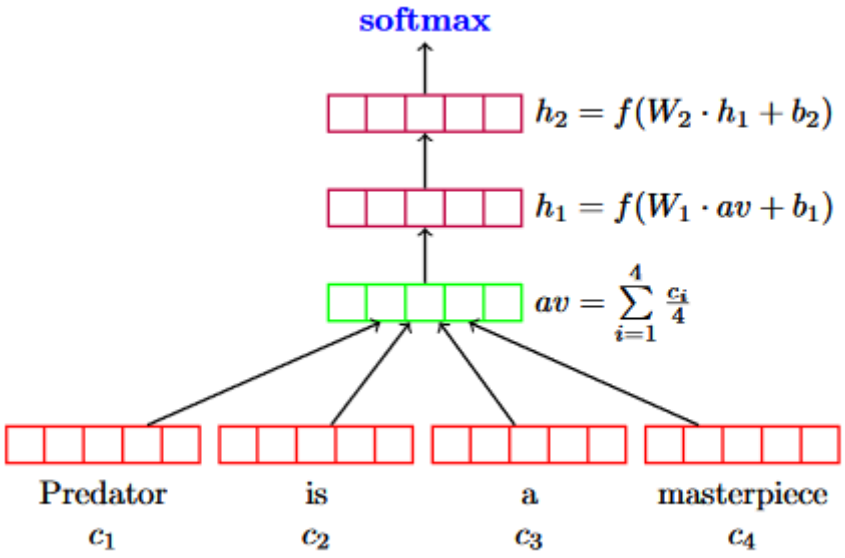
แบบจำลองการถดถอยโลจิสติก จะให้ความน่าจะเป็นสูงสุดในป้ายกำกับที่ถูกต้อง (y) เมื่อสามารถหาค่าน้ำหนัก (w) ที่ทำให้ลอการิทึมของความเป็นไปได้ลบ (Negative Log-likelihood) นี้ต่ำที่สุด ซึ่งสามารถทำได้ผ่านกระบวนการหาค่าน้ำหนักที่เหมาะสมที่สุด (Optimization) โดยวิธีที่นิยมใช้คืออัลกอริทึมการเคลื่อนลงตามความชัน (Gradient

Descent Algorithm) เช่น สโตแคสติกเกรเดียนต์เดสเซนท์ (Stochastic Gradient Descent) และแบตช์เกรเดียนต์เดสเซนท์ (Batch Gradient Descent) ซึ่งล้วนมีหน้าที่ในการปรับหาค่าน้ำหนักที่ทำให้ค่าสูญเสีย (Loss) ของแบบจำลองต่ำที่สุด โดยสโตแคสติกเกรเดียนต์เดสเซนท์จะคำนวณและปรับค่าน้ำหนักในแต่ละข้อมูล ในขณะที่แบตช์เกรเดียนต์เดสเซนท์จะคำนวณค่าและปรับค่าน้ำหนักของข้อมูลทั้งหมดพร้อมกันก่อนที่จะดำเนินการปรับค่าน้ำหนักต่อไป

2.1.3 Deep Averaging network (DAN)

Deep Averaging Network (DAN) (Iyyer, Manjunatha et al. 2015) เป็นแบบจำลองเชิงลึกที่ถูกออกแบบมาเพื่อการ ประมวลผลภาษาธรรมชาติ (Natural Language Processing) ซึ่งมีความสามารถใน การเรียนรู้ความหมายของประโยคหรือข้อความต่างๆ โดยใช้หลักการเฉลี่ยข้อมูลจาก เวกเตอร์คำ (word vectors) ที่สร้างจากแบบจำลองที่ถูกฝึกสอนล่วงหน้า เช่น Word 2Vec หรือ GloVe และประมวลผลเฉลี่ยผ่านเลเยอร์ลึก (deep layers) ของแบบจำ ลองเพื่อให้ได้เวกเตอร์ประเมินความหมายของประโยคที่มีความแม่นยำ สูงขึ้น

หลักการทำงานของ Deep Averaging Network (DAN) สามารถอธิบายได้ตามขั้นตอนต่อไปนี้ (ดังภาพที่ 1):




*รูปภาพ 1 Deep Averaging Network ที่มา (Iyyer, Manjunatha et al. 2015)*

*2.1.3.1 Word Embedding*

ในขั้นตอนแรกเนื้อหาของประโยคหรือข้อความจะถูกแปลงเป็น เวกเตอร์คำ ซึ่งสามารถทำได้โดยใช้แบบจำลอง Word Embedding เช่น Word2Vec หรือ GloVe ที่ได้รับการฝึกสอนล่วงหน้า โดยแต่ละคำจะถูกแทนที่ด้วยเวกเตอร์ที่เกี่ยวข้องกับความหมายของคำนั้น ๆ

*2.1.3.2 Averaging*

เวกเตอร์ของคำทั้งหมดในประโยคหรือข้อความจะถูกเฉลี่ยเพื่อสร้าง เวกเตอร์เฉพาะของประโยค ได้แก่ ผลรวมของเวกเตอร์คำทั้งหมดจะถูก แบ่งด้วย จำนวนคำทั้งหมดในประโยค

*2.1.3.3 Deep Layers*

เวกเตอร์ประโยคที่ได้จากขั้นตอนก่อนหน้าจะถูกนำเข้าสู่เลเยอร์ลึกของแบบจำลอง เลเยอร์ลึกประกอบด้วยเลเยอร์ประสาทเทียม (neural layers) ที่ประกอบไปด้วยเซลล์ประสาทที่มีการเชื่อมโยงกันอย่างชนิด แบบเต็มเชิง (fully connected) โดยแต่ละเลเยอร์จะมีการใช้ฟังก์ชัน เปิดสังวาดีเอฟเนิร์ (Rectified Linear Unit, ReLU) เพื่อเพิ่มประสิทธิ ภาพ ในการเรียนรู้และสกัดลักษณะเชิงลึกของประโยค

*2.1.3.4 Classification*

ในขั้นตอนสุดท้ายเวกเตอร์ที่ได้จากเลเยอร์ลึกจะถูกนำไปใช้ในการ จำแนกหรือทำนายประเภทของประโยคโดยใช้เลเยอร์ที่มีการเชื่อม โยงแบบเต็มเชิง (fully connected layer) ซึ่งส่งผลลัพธ์ออกมาเป็นเวกเตอร์คลาสที่แสดงประเภทที่แบบจำลองได้ทำนายขึ้นมา

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

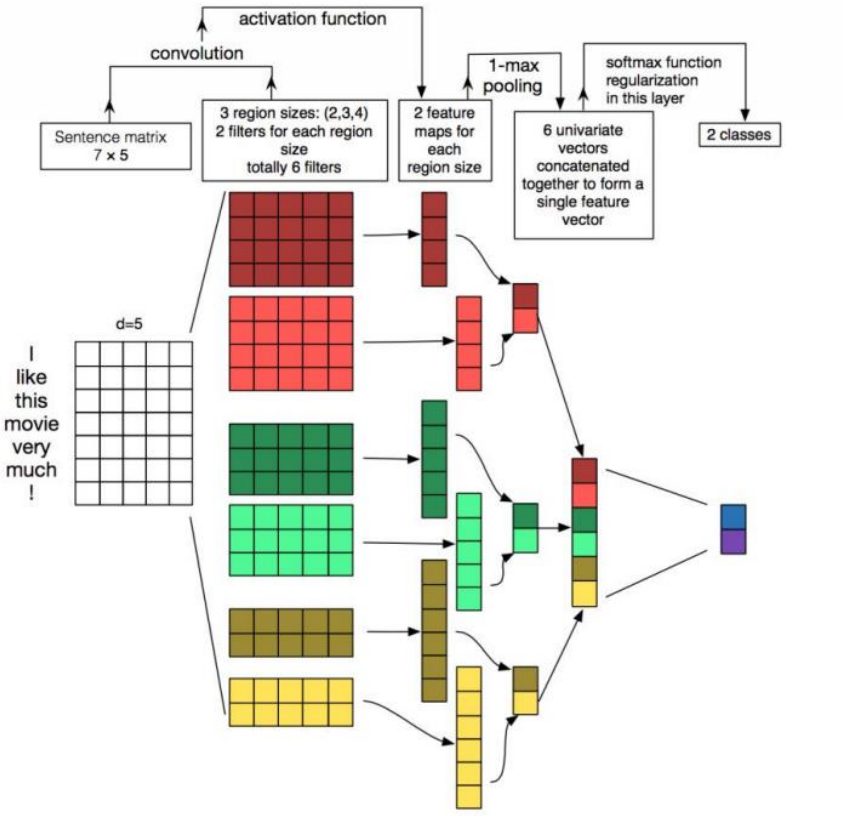
**2.1.4 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network)**

โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) (LeCun, Bengio et al. 2015) เป็นแบบจำลองการเรียนรู้เชิงลึกที่ใช้ในการประมวลผลข้อมูลที่มีลักษณะเป็นตารางหรือภาพ โดยเฉพาะอย่างยิ่งในงานประมวลผลภาพและการจำแนกภาพ (image classification) ซึ่ง CNN ได้รับความนิยมและนำมาใช้อย่างแพร่หลายในหลากหลายงานด้วยประสิทธิภาพและประสิทธิผลที่สูง

CNN ถูกออกแบบมาเพื่อจำลองการประมวลผลของสมองมนุษย์ในการรับรู้และจำแนกภาพ โครงสร้างพื้นฐานของ CNN ประกอบด้วยชั้นคอนโวลูชัน (Convolutional Layer) และชั้นพูลลิ่ง (Pooling Layer) ที่สลับกันเพื่อสกัดคุณลักษณะหรือลักษณะที่น่าสนใจ

ของภาพออกมาให้เป็นเอกลักษณ์ (features) และชั้นแบบเพดาน (Fully Connected Layer) ที่ใช้ในการจำแนกหรือคาดเดาผลลัพธ์

โครงสร้างหลักของ CNN คือการใช้ตัวกรอง (filter) หรือเคอร์เนล (kernel) ในการทำคอนโวลูชันบนภาพ ซึ่งตัวกรองจะเป็นแมทริกซ์ขนาดเล็กที่ทำการคูณสมาชิกกับพิกัดหนึ่งๆของภาพ แล้วหาผลรวมเพื่อสร้างเอกลักษณ์ใหม่ (feature map) ของภาพ โดยที่ตำแหน่งของเอกลักษณ์จะขยายตามการเคลื่อนที่ของตัวกรองในการคอนโวลูชัน นอกจากนี้การใช้ชั้นพูลลิ่ง (pooling layer) จะช่วยลดขนาดของเอกลักษณ์และสกัดคุณลักษณะสำคัญออกมาเพื่อลดความซับซ้อนและความจำเป็นในการคำนวณ โดยการฝึกฝน (training) ของ CNN นั้นใช้กระบวนการถดถอยแบบหลายชั้น (backpropagation) เพื่อปรับค่าน้ำหนักและพารามิเตอร์ในแบบจำลองเพื่อให้ได้ผลลัพธ์ที่ใกล้เคียงกับผลลัพธ์ที่ถูกต้อง ซึ่งการปรับค่าน้ำหนักและพารามิเตอร์จะใช้อัลกอริทึมการเรียนรู้เชิงความลึกเช่นอัลกอริทึมการเรียนรู้ออกแบบใหม่ (Adam) หรืออัลกอริทึมการเรียนรู้มีเสถียรภาพ (SGD with Momentum) เพื่อปรับค่าน้ำหนักให้ลดค่าสูญเสีย (loss) อย่างมากที่สุด (ดังตัวอย่างภาพที่ 2)




*รูปภาพ 2 กระบวนการใช้แบบจำลอง CNN กับข้อความ (Zhang and Wallace 2015)*

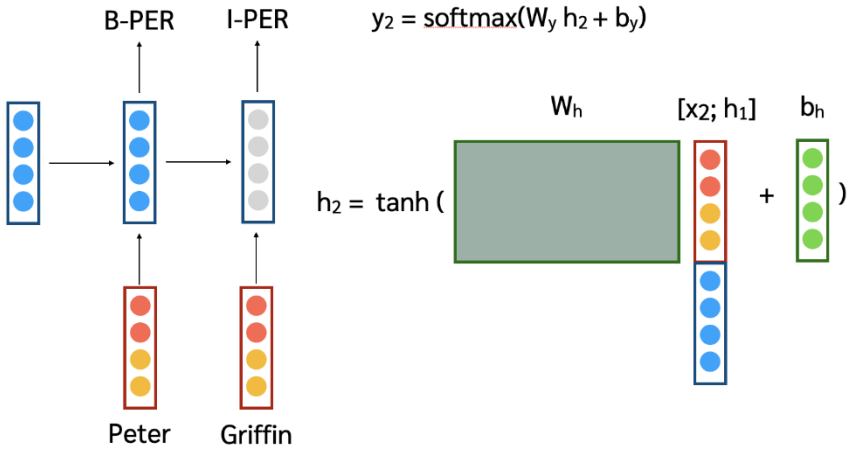
ด้วยการคัดลอกและรวมกันของชั้นคอนโวลูชันและชั้นพูลลิ่งต่างๆ ของ CNN สามารถจับคุณลักษณะที่ซับซ้อนของภาพได้อย่างมีประสิทธิภาพ และสามารถจัดลำดับหรือจำแนกภาพในกลุ่มหรือประเภทที่ต่างกันได้. โดย CNN ได้รับความนิยมอย่างมากในงานที่เกี่ยวข้องกับการประมวลผลภาพ เช่น การจดจำใบหน้า (face recognition), การตรวจจับวัตถุ (object detection), การประมวลผลภาพทางการแพทย์ (medical image processing), และการประมวลผลภาพทางด้านสื่อสารมวลชน (media processing) ซึ่งมีการประยุกต์ใช้ที่หลากหลายและมีผลลัพธ์ที่ดีในงานนี้

**2.1.5 โครงข่ายประสาทเทียมแบบวงวน (Recurrent Neural Network)**

โครงข่ายประสาทเทียมแบบวงวน (Recurrent Neural Network, RNN) เป็นแบบจำลองที่สามารถประมวลผลข้อมูลที่มีการเรียงตัวเป็นลำดับได้อย่างมีประสิทธิภาพ โดยเฉพาะอย่างยิ่งในการประมวลผลข้อมูลที่เป็นรูปประโยค โดยไม่จำเป็นต้องกำหนดจำนวนคำในประโยคล่วงหน้า ทำให้ RNN เป็นแบบจำลองที่มีประโยชน์อย่างมากในงานที่เกี่ยวข้องกับการประมวลผลภาษาธรรมชาติ โดยที่ผลลัพธ์ที่ได้จากแบบจำลอง RNN สามารถนำไปใช้ในการจำแนกข้อความ (Text Classification) หรือนำมาใช้ในการกำกับชื่อหรือชนิดของหน่วยภาษาในข้อความ (Sequence Labeling) เช่น การรู้จำชื่อเฉพาะ (Named-entity Recognition) เป็นต้น ซึ่ง RNN มีส่วนประกอบสำคัญที่ช่วยในการประมวลผลภาษาธรรมชาติ คือ เวกเตอร์ของคำที่ผ่านการเรียนรู้โดยใช้ชุดข้อมูลอื่นที่ได้รับการเรียนรู้ไว้ล่วงหน้า (Pretrained Word Embedding) เป็นตัวอย่างของแบบจำลองการเรียนรู้เชิงลึกที่ได้กล่าวถึงข้างต้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

แบบจำลอง Vanilla RNN เป็นแบบจำลองพื้นฐานที่มีโครงสร้างไม่ซับซ้อน ของแบบจำลองแบบ RNN เรียกว่า Vanilla Recurrent Neural Network (Vanilla RNN) วิธีการทำงานของแบบจำลองดังกล่าวคือการนำ Word Embedding (x) มาต่อกับเวกเตอร์กระตุ้นแอบแฝง (Hidden Activation Vector: h) แล้วจึงนำมาคูณกับเมทริกซ์ค่าน้ำหนัก (Wh) และบวกด้วยเวกเตอร์ค่าความเอนเอียง (bh) ก่อนนำเข้าไปประมวลผลในฟังก์ชันกระตุ้นที่ไม่เป็นเชิงเส้นตรง (Non-linear Activation Function) เช่น ฟังก์ชันไฮเพอร์โบลิกแทนเจนต์ (Hyperbolic Tangent: tanh) ดังรูปที่ 3 โดยมีลักษณะที่คล้ายกับโครงข่ายประสาทเทียมแบบแพร่ไปข้างหน้า สุดท้ายจะได้เวกเตอร์กระตุ้นแอบแฝงของคำที่ตำแหน่งนั้นซึ่งได้ซึมซับข้อมูลจากตำแหน่งที่ผ่านมาและข้อมูลจากคำที่อยู่ในตำแหน่งปัจจุบัน ซึ่งจะถูกนำไปใช้คำนวณเวกเตอร์กระตุ้นแอบแฝงในตำแหน่งถัดไป หรือนำไปใช้เป็นเวกเตอร์ค่าสัญญาณในการทำนายป้ายกำกับได้

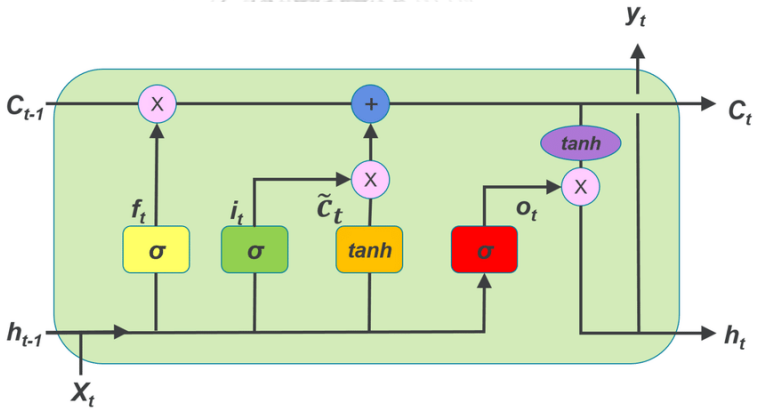



*รูปภาพ 3 แผนภาพการทำงานของแบบจำลอง Vanilla RNN*

ปัญหาที่พบบ่อยในการใช้แบบจำลอง Vanilla RNN คือปัญหาของการสอดคล้องกับการปรับค่าเกรเดียนที่มีขนาดต่างกัน ซึ่งอาจทำให้แบบจำลองไม่สามารถเรียนรู้ได้อย่างมีประสิทธิภาพสูงสุด ปัญหานี้เกิดจากการคำนวณหาค่าเกรเดียนที่เหมาะสมที่สุดในกระบวนการปรับค่าน้ำหนัก ซึ่งเป็นกระบวนการที่มีการคูณกันหลายขั้นตอน ทำให้ตัวเลขที่มีค่าน้อยมีค่าลดลงไปเรื่อย ๆ จนไม่มีผลต่อการเรียนรู้ของแบบจำลอง หรือตัวเลขที่มีค่ามากก็จะถูกขยายเพิ่มขึ้นจนทำให้ค่าน้ำหนักมีค่าสูงจนเกินไป (Exploding Gradient) ในการใช้แบบจำลอง RNN จึงมีการพัฒนาแบบจำลอง RNN ชนิดอื่นเช่น LSTM (Long-short Term Memory) หรือ GRU (Gated Recurrent Unit) ในการควบคุมค่าเกรเดียนให้อยู่ในช่วงที่เหมาะสม ซึ่งทำให้แบบจำลองที่มีขนาดต่าง ๆ สามารถเรียนรู้เพื่อหาขนาดของแบบจำลองที่มีประสิทธิภาพสูงสุดได้

โครงข่ายประสาทเทียมแบบวงวนเป็นแบบจำลองที่มีความสามารถสูงเพราะสามารถเก็บความหมายจากข้อความได้โดยไม่ขึ้นอยู่กับความกว้างของฟิลเตอร์เหมือนกับแบบจำลอง CNN ในกรณีที่ฟิลเตอร์ของ CNN มีขนาด 3 คำ แบบจำลองจะไม่สามารถเก็บความหมายจากกลุ่มคำที่รวมกันแล้วมีความหมายใหม่ขนาด 4 คำขึ้นไปได้ ในทางตรงข้ามแบบจำลองแบบ LSTM เป็นแบบจำลองโครงข่ายประสาทเทียมแบบวงวนแบบหนึ่งที่สามารถประมวลผลข้อความโดยไม่จำกัดจำนวนคำด้วยการใช้เมทริกซ์ค่าน้ำหนักเดิมสำหรับทุกคำ ทุกตำแหน่งของประโยค บางครั้งแบบจำลอง RNN แบบ GRU จะถูกนำมาใช้แทนแบบจำลองแบบ LSTM เนื่องจากสามารถประมวลผลได้เร็วกว่า ทั้งนี้ทั้งสองแบบจำลองนั้นมีส่วนประกอบที่สำคัญคือ (1) เซลล์จดจำ (Memory Cell) และ (2) โครงข่ายประตูควบคุมการไหลของข้อมูล (Gating Network)

เซลล์จดจำ (c ในรูปที่ 4) ในแบบจำลอง LSTM เป็นส่วนที่เก็บความหมายที่ได้มาจากคำที่ประมวลผลมาแล้วเพื่อนำมาเป็นบริบทในการประมวลผลคำถัดไป อีกหนึ่งข้อดีของแบบจำลอง LSTM คือการควบคุมการไหลของข้อมูล (Gating) ซึ่งได้รับแรงบันดาลใจมาจากการต่อแผงวงจร ประตู (Gate) เป็นส่วนที่ปิดเปิดเพื่อเลือกว่าจะนำส่วนใดจากคำใหม่เข้ามาเป็นส่วนหนึ่งของเซลล์จดจำและเวกเตอร์กระตุ้นแอบแฝง (h ในรูปที่ 2.11) ทำให้แบบจำลองนี้มีความหลากหลายมากขึ้น แม้แบบจำลองจะสามารถถูกออกแบบให้มีรายละเอียดต่างกันได้ แต่ส่วนที่สำคัญที่สุดคือการมีประตูเลือกลืม (Forget Gate) เพื่อเลือกส่วนที่จะนำเข้ามาเพิ่มเติม และผสมกับความหมายที่ประมวลผลมาแล้ว แบบจำลอง LSTM จะประมวลผลคำจากซ้ายไปขวา เมื่อประมวลผลครบทุกคำแล้วจะได้เวกเตอร์กระตุ้นแอบแฝงตำแหน่งล่าสุด ซึ่งเป็นเวกเตอร์เก็บข้อมูลระดับประโยค (Sentence Embedding) ที่จะถูกนำไปใช้ในการจำแนกข้อความต่อไป แบบจำลอง LSTM คล้ายกับโครงข่ายประสาทเทียมเฉลี่ยเชิงลึก (Deep Averaging: DAN) และโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) ในเรื่องของการเก็บข้อมูลระดับประโยค




*รูปภาพ 4 แผนภาพแบบจำลอง LSTM (Ismail, Wood et al. 2018)*

$$\mathbf{i}_t = \sigma\left(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_i\right)$$
$$\mathbf{f}_t = \sigma\left(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_f\right)$$
$$\mathbf{o}_t = \sigma\left(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_o\right)$$
$$\tilde{\mathbf{c}}_t = \tanh\left(\mathbf{W}_{xc}\mathbf{x}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c\right)$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t$$
$$\mathbf{h}_t = \tanh\left(\mathbf{o}_t \odot \mathbf{c}_t\right)$$

*รูปภาพ 5 LSTM มีการประมวลผลคำแต่ละคำ โดยใช้สมการดังนี้ (Ismail, Wood et al. 2018)*

$\sigma$ คือ ฟังก์ชันซิกมอยด์โลจิสติก (Logistic Sigmoid Function)

$\odot$ คือ ผลคูณแบบ Hadamard (การหาผลคูณแบบมิติต่อมิติ)

Wi Wf Wo Wc คือ เมทริกซ์ค่าน้ำหนักแบบวงวน (Recurrent Weight Matrix) ของประตูเลือกรับ (Input Gate) ประตูเลือกลืม (Forget Gate) ประตูเลือกส่ง (Output Gate) และประตูเลือกข้อมูล (Input Modulate Gate) ตามลำดับ

bi bf bo bc คือ ค่าความเอนเอียงของประตูเลือกรับ (Input Gate) ประตูเลือกลืม (Forget Gate) ประตูเลือกส่ง (Output Gate) และประตูเลือกข้อมูล (Input Modulate Gate) ตามลำดับ

ht คือ เวกเตอร์กระตุ้นแอบแฝงที่คำนวณมาจากคำที่ t ของข้อความนี้

it, ft, ot, c~t ct คือประตูเลือกรับ (Input Gate) ประตูเลือกลืม (Forget Gate) ประตูเลือกส่ง (Output Gate) ประตูเลือกข้อมูล (Input Modulate Gate) และเซลล์จดจำ (Memory Cell) ตามลำดับ ซึ่งเป็นเวกเตอร์ที่มีขนาดเท่ากันทั้งหมด

**2.1.6 แบบจำลองตัวเข้ารหัสแบบสองทิศทางในรูปแบบตัวแปลง (Bidirectional Encoder Representations from Transformers: BERT)**

แบบจำลองตัวเข้ารหัสแบบสองทิศทางในรูปแบบตัวแปลง (Bidirectional Encoder Representations from Transformers: BERT) มีโครงสร้างที่ประกอบด้วย Transformer 12 ชั้นที่เชื่อมต่อกัน แต่ละชั้นเป็นโมดูล Transformer ที่มีลักษณะเหมือนกัน โดยรับข้อมูล (input) เป็น Word Embedding ของแต่ละคำในข้อความ และส่งข้อมูลออก (output) เป็น Word Embedding ที่ผ่านการประมวลผลเพื่อวิเคราะห์บริบทของคำอื่นๆ ในข้อความด้วยกัน การคำนวณหา Word Embedding ในแบบจำลอง BERT ใช้กลไกที่เรียกว่า Self-attention ซึ่งใช้ในการหา Word Embedding ของแต่ละคำโดยการเฉลี่ยแบบถ่วงน้ำหนักของคำทั้งหมดที่อยู่ในประโยคนั้น ค่าน้ำหนักที่ให้กับ Word Embedding แต่ละคำเรียกว่า Attention ซึ่งคำนวณโดยใช้ค่า Word Embedding ทั้งหมด 3 ชุด คือ:

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

**Query:** Word Embedding ของคำที่ต้องการวิเคราะห์และคำนวณหา Embedding เพื่อนำออกมาเป็นข้อมูลส่งออก (output)

**Key:** Word Embedding ของคำอื่นๆ ที่อยู่ในข้อความ

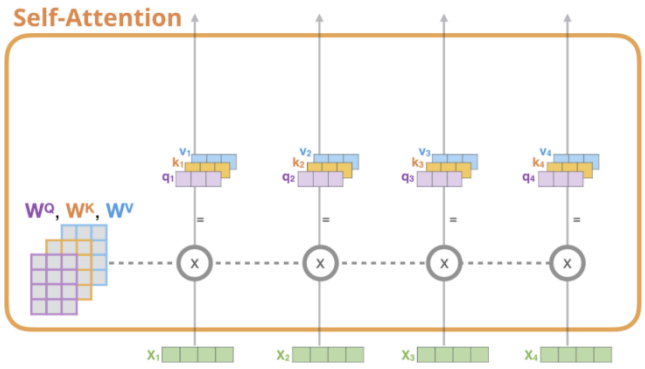
**Value:** Word Embedding ที่ใช้ในการคำนวณค่าเฉลี่ยแบบถ่วงน้ำหนัก เพื่อนำออกมาเป็นข้อมูลส่งออก (output)

ใน Transformer แต่ละชั้น Embedding ของคำทุกคำในรายการคำศัพท์ของแบบจำลองไม่ถูกสร้างเป็น Embedding ที่เจาะจงสำหรับแต่ละคำ แต่จะมีการ คำนวณ

เวกเตอร์ Query (q), Key (k), และ Value (v) โดยใช้การแปลงเชิงเส้น (Linear Transformation) กล่าวคือ Word Embedding ของคำที่ต้องการหา เวกเตอร์ Query จะถูกคูณกับเมทริกซ์ค่าน้ำหนักสำหรับ Query ($W^Q$) ส่วนการคำนวณเวกเตอร์ Key และเวกเตอร์ Value นั้นใช้กระบวนการแปลงเชิงเส้น โดยการคูณด้วยเมทริกซ์ค่าน้ำหนักสำหรับ Key ($W^K$) และเมทริกซ์ค่าน้ำหนักสำหรับ Value ($W^V$) ตามลำดับ

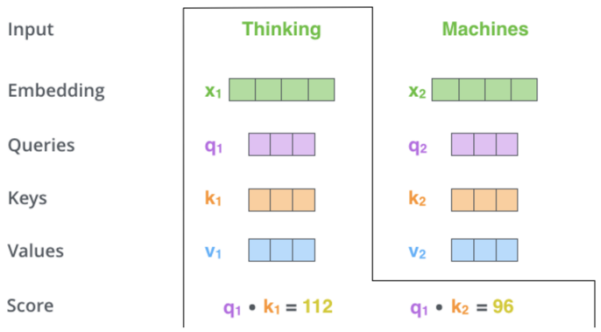
$$q_i = W^Q x_i \quad (2)$$
$$k_i = W^k x_i \quad (3)$$
$$v_i = W^v x_i \quad (4)$$

โดยที่ xi qi ki vi เป็น Word Embedding, เวกเตอร์ Query, เวกเตอร์ Key และเวกเตอร์ Value ตามลำดับ



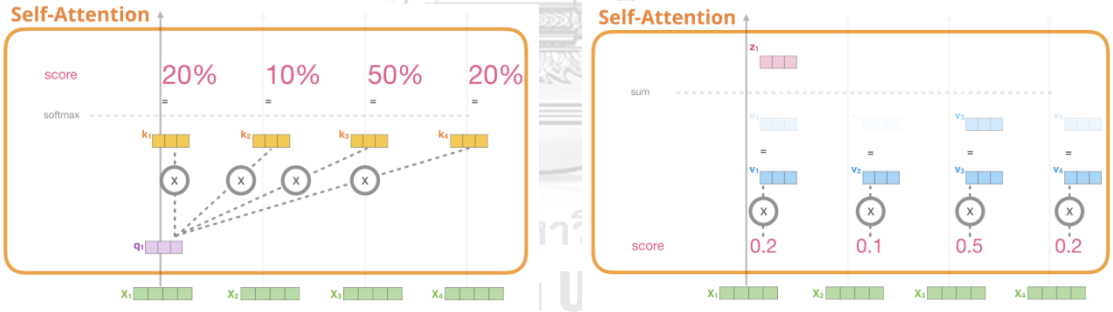

*รูปภาพ 6 แผนภาพการคำนวณเวกเตอร์ Query เวกเตอร์ Key และเวกเตอร์ Value ของแต่ละคำที่ปรากฏในข้อความ (Alammar 2018)*

หลังจากที่ได้เวกเตอร์ Query, เวกเตอร์ Key และเวกเตอร์ Value ของทุกคำในข้อความแล้ว การคำนวณ Attention ของคำแต่ละคำที่อยู่ในข้อความได้มาจากการหาผลคูณเชิงสเกลาร์ (Dot Product) ของเวกเตอร์ Query และเวกเตอร์ Key ดังรูปที่ 7




*รูปภาพ 7 แผนภาพการคำนวณ attention ของคำที่อยู่ใกล้เคียง (Alammar 2018)*

ยกตัวอย่างประโยคที่มีคำว่า Thinking Machines เป็นจำนวน 2 คำ ขั้นตอนแรกเริ่มจากการคำนวณหา Embedding ของคำว่า Thinking ที่อยู่ในประโยคนี้โดยการหาผลคูณเชิงสเกลาร์ของเวกเตอร์ Query ของคำว่า Thinking และเวกเตอร์ Key ของคำว่า Machines แล้วจะได้ผลลัพธ์ออกมาเป็นค่า Attention ของคำว่า Thinking ที่ให้กับคำว่า Machines จากนั้นหาผลคูณเชิงสเกลาร์ของเวกเตอร์ query ของคำว่า Thinking และเวกเตอร์ Key ของคำว่า Thinking แล้วจะได้ผลลัพธ์ออกมาเป็นค่า Attention ของคำว่า Thinking ที่ให้กับคำว่า Thinking (ซึ่งเป็นที่มาของคำว่า Self-attention) เมื่อกระบวนการนี้เสร็จสิ้น ขั้นตอนต่อไปคือการคำนวณหา Embedding ตัวใหม่ของคำว่า Thinking โดยนำ เวกเตอร์ value ของคำว่า Thinking มาเฉลี่ยกับเวกเตอร์ Value ของคำว่า Machines โดยถ่วงการหาค่าเฉลี่ยด้วยค่า Attention ที่คำว่า Thinking ให้กับตัวเอง และให้กับคำว่า Machines ตามลำดับ กลไก Self-attention จึงเปรียบเสมือนการวิเคราะห์ความหมายของคำแต่ละคำ โดยต้องนำคำที่อยู่รอบ ๆ มาวิเคราะห์ร่วม และไม่ใช่ทุกคำจะมีบทบาทในการแปลความหมายเท่ากันหมด กลไกนี้จึงเป็นกลไกที่มีประสิทธิภาพมากที่ทำให้เครื่องได้เรียนรู้ว่าคำใดควรวิเคราะห์ร่วมกับคำใดบ้าง จึงจะเข้าใจความหมายของคำ ๆ นั้น และความหมายโดยรวมของทั้งประโยคได้




*รูปภาพ 8 แผนภาพการประกอบสร้าง Word Embedding ใหม่จาก Attention และเวกเตอร์ Value (Alammar 2018)*

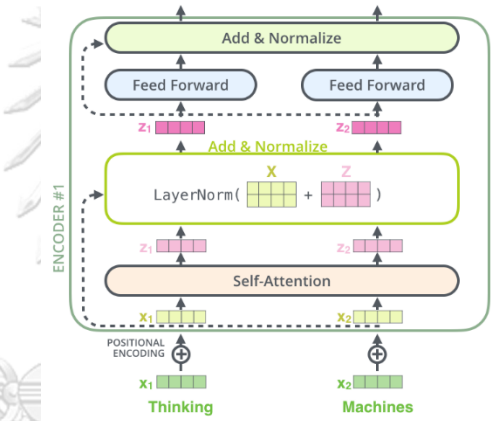
การคำนวณค่า Attention สามารถเขียนเป็นสูตรได้ดังนี้: สมมติให้ xi เป็น Word Embedding ของคำที่ i ที่ถูกป้อนผ่านกระบวนการ Self-attention แล้ว และ $x_j$ เป็น Word Embedding ของคำที่อยู่รอบ ๆ

$$a_j = softmax(\Sigma_j q_i \cdot k_j) \quad (5)$$

$$z_i = \Sigma_j v_j \cdot a_j \quad (6)$$

โดยที่ $a_j$ คือค่า Attention ที่ให้กับคำที่ j ในข้อความ และ softmax คือ ฟังก์ชัน softmax ที่คำนวณลงไปบนคำแต่ละคำในประโยค ส่วน $z_i$ คือ Word Embedding ตัวใหม่ที่ได้จากการคำนวณจากเวกเตอร์ Value ($v_j$) ที่อยู่รอบ ๆ

สถาปัตยกรรม Transformer ในแบบจำลอง BERT ใช้การทำ Self-attention โดยตัวแบบจำลองนี้จะใช้การคำนวณ Self-attention กับทุกคำในข้อความ เพื่อให้ได้ Word Embedding ชุดใหม่สำหรับทุกคำ เพื่อนำไปประมวลผลต่อและส่งให้ชั้น Transformer อื่น ๆ ได้ประมวลผลต่อด้วยสมการเดียวกัน อย่างไรก็ตาม ชุดพารามิเตอร์ Query, Key, และ Value จะมีค่าที่แตกต่างกันในแต่ละชั้น และจะถูกปรับแต่งให้เป็นค่าที่สามารถให้ Word Embedding ที่เป็นประโยชน์โดยการเรียนรู้จากข้อมูลชุดอื่นมาก่อน (pretraining) นอกจาก Self-attention แล้ว Transformer ยังมีส่วนประกอบอื่น ๆ เข้ามาช่วยในการแปลงค่าด้วย




*รูปภาพ 9 แผนภาพการเข้ารหัสของตำแหน่งคำ และการบวกและการปัดให้เป็นค่าปกติ (Alammar 2018)*

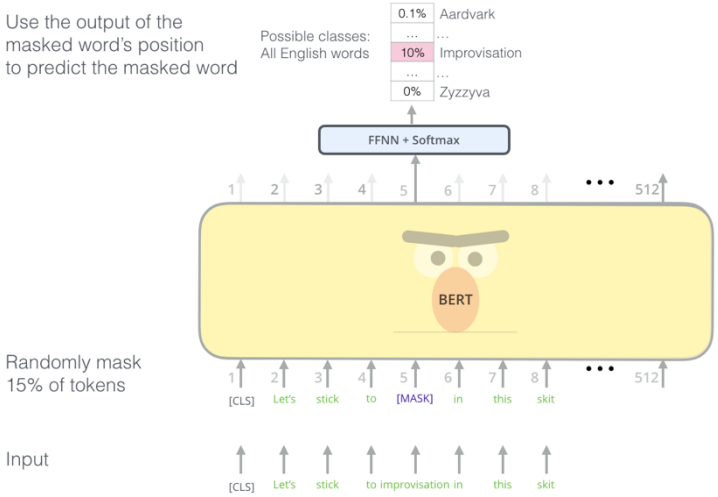
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Positional Encoding เป็นกระบวนการที่ทำให้ค่าของ Embedding เปลี่ยนแปลงตามตำแหน่งที่เกิดขึ้น โดย Self-attention ไม่คำนึงถึงตำแหน่งและลำดับของคำ ซึ่งส่งผลให้ Positional Encoding เป็นสิ่งสำคัญในการเพิ่มความเข้าใจให้กับแบบจำลอง ผลลัพธ์ของ Self-attention ไม่ได้เป็นผลลัพธ์ส่งออกโดยตรง แต่จะถูกบวกกับข้อมูลรับเข้าและปรับค่าให้เป็นค่าปกติ จากนั้นผลลัพธ์จะถูกส่งให้กับโครงข่ายประสาทเทียมแบบแพร่ไปข้างหน้า หลังจากนั้นจะมีกระบวนการ add and normalization อีกครั้งก่อนที่จะได้ผลลัพธ์สุดท้ายเป็นผลลัพธ์ส่งออกของ Transformer

แบบจำลอง Transformer ประกอบด้วยพารามิเตอร์จำนวนมากในแต่ละชั้น แบบจำลอง BERT ใช้ Transformer 9-12 ชั้น การฝึกแบบจำลองนี้ต้องใช้พารามิเตอร์จำนวนมาก โดยกระบวนการ Pretraining เป็นกระบวนการที่ฝึกพารามิเตอร์ทั้งหมดใน

แบบจำลองโดยให้เรียนรู้จากข้อมูลชุดใหญ่ ๆ ของข้อความ จะทำให้แบบจำลองมีความรู้เกี่ยวกับภาษามาก่อนที่จะนำไปใช้ในงานอื่น ๆ กระบวนการ pretraining นี้คล้ายกับการฝังคำ (Word Embedding) แต่ BERT นั้นฝึกเรียนความหมายของคำและบริบทในข้อมูลในเวลาเดียวกัน เรียกว่าการฝังคำแบบใช้บริบทประกอบ (Contextualized Word Embeddings) ผลลัพธ์ที่ได้คือ Contextualized Word Embedding และ Sentence Embedding ของทั้งประโยค ซึ่งสามารถนำไปใช้ในงานอื่น ๆ

Masked Language Modeling เป็นงานสำคัญในการฝึกแบบจำลอง Transformer เพื่อเพิ่มความรู้ให้กับ Transformer เกี่ยวกับคำศัพท์ที่เจอในคลังข้อมูล รวมถึงความสามารถในการประมวลผลความหมายของคำศัพท์ในบริบทของประโยคที่ไม่เคยพบมาก่อน การทำ Masked Language Modeling คล้ายกับวิธีการ word2vec (Continuous bag-of-word) โดยสุ่มคำในข้อความแล้วแทนคำด้วยสัญลักษณ์ [MASK] แล้วให้ป้อนข้อความเข้าสู่ Transformer เพื่อคำนวณ Word Embedding ที่นำไปเป็นข้อมูลรับเข้าให้กับ Transformer ต่อไป แบบจำลอง BERT ใช้ Transformer 9 ชั้นสำหรับ BERT base และ 12 ชั้นสำหรับ BERT ชุดใหญ่ โดยผลลัพธ์คือ Word Embedding ตัวสุดท้ายที่เก็บความหมายของคำที่ถูกวิเคราะห์ในบริบทของประโยค ซึ่งนำไปทำนายคำที่ถูกปิดไว้

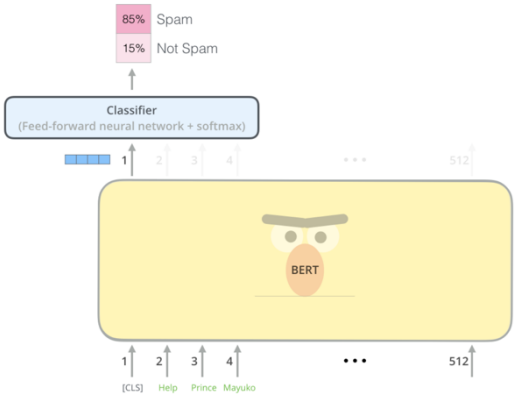



*รูปภาพ 10 แผนภาพกระบวนการ Pretraining แบบจำลอง BERT โดยการใช้ Masked Language Modeling Task (Alammar 2018)*

ในกระบวนการ Pretraining ของแบบจำลอง Transformer เช่น BERT นั้น มักจะใช้ข้อมูลดิบที่มีขนาดอย่างน้อยพันล้านคำเพื่อฝึกแบบจำลอง โดยมีจำนวนพารามิเตอร์ทั้งหมดในแบบจำลอง BERT base อยู่ที่ 110 ล้านตัว และในแบบจำลอง BERT ชุดใหญ่

จำนวนพารามิเตอร์ทั้งหมดอยู่ที่ 340 ล้านตัว ในกระบวนการ Pretraining นี้ จำเป็นต้องใช้เวลาประมาณ 12 วันในการฝึกแบบจำลอง BERT โดยใช้เครื่องคอมพิวเตอร์ที่มี GPU รุ่น V100 เพื่อทำการคำนวณ ซึ่งเป็นเครื่องคอมพิวเตอร์ที่มีพลังการคำนวณสูงและมีราคาแพง ในกรณีนี้ใช้จำนวนเครื่องคอมพิวเตอร์ 8 เครื่องเพื่อฝึกแบบจำลองในเวลาที่เหมาะสม

อย่างไรก็ตาม ข้อดีของ BERT คือ เมื่อแบบจำลองผ่านกระบวนการ Pretraining เรียบร้อยแล้ว สามารถนำแบบจำลองไปประยุกต์ใช้กับงานทางภาษาต่าง ๆ ได้มากมาย เนื่องจากความรู้ที่เกิดขึ้นจากกระบวนการนี้สามารถถูกถ่ายทอดไปใช้ในงานอื่น ๆ ได้ ยกตัวอย่างเช่น สามารถทำกระบวนการปรับจูน (Fine-tuning) แบบจำลอง BERT ให้เข้ากับงานที่ต้องการการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยให้แบบจำลองมีข้อมูลตัวอย่างที่ชัดเจนเกี่ยวกับข้อมูลรับเข้าและข้อมูลส่งออก โดยอาจจะมีจำนวนข้อมูลที่ไม่มากและไม่เหมาะสำหรับการใช้แบบจำลองที่ไม่ผ่านกระบวนการ Pretraining และการถ่ายทอดความรู้




*รูปภาพ 11 แผนภาพกระบวนการ Fine-tuning เพื่อถ่ายทอดความรู้จากแบบจำลองที่ถูกเทรนมาก่อนส่วนหนึ่ง (Alammar 2018)*

ในกระบวนการ Fine-tuning จะนำแบบจำลอง BERT ที่ผ่านกระบวนการ Pretraining มาปรับแต่งให้เหมาะสมกับงานที่ต้องการ โดยมีข้อมูลตัวอย่างที่ชัดเจนเป็นแนวทางในการปรับแต่ง แบบจำลอง BERT ที่ถูกฝึกมาล่วงหน้าผ่านข้อมูลหลายพันล้านคำจะถูกปรับแต่งเพื่อให้เข้ากับบริบทของงานนั้น ๆ โดยจะเรียกกระบวนการนี้ว่า Fine-tuning ซึ่งเป็นขั้นตอนที่ทำให้แบบจำลองเข้าใจความหมายของคำศัพท์ทั้งแบบไม่มีบริบทและแบบมีบริบท รวมถึงกฎไวยากรณ์ต่าง ๆ ที่ส่งผลต่อการตีความหมายของคำและประโยค จากนั้นสามารถนำแบบจำลองที่มีความรู้ทางภาษาเพื่อประยุกต์ใช้ในงานต่าง ๆ ที่มีความเฉพาะด้านมากขึ้นและปรับความรู้ทางภาษาไปด้วยในเวลาเดียวกัน โดยการปรับจูนนี้จะทำ

ให้แบบจำลอง BERT เป็นตัวสกัดค่าสัญญาณ (Feature Extractor) ที่สามารถใช้ในงานที่ต้องการการประมวลผลภาษาธรรมชาติได้

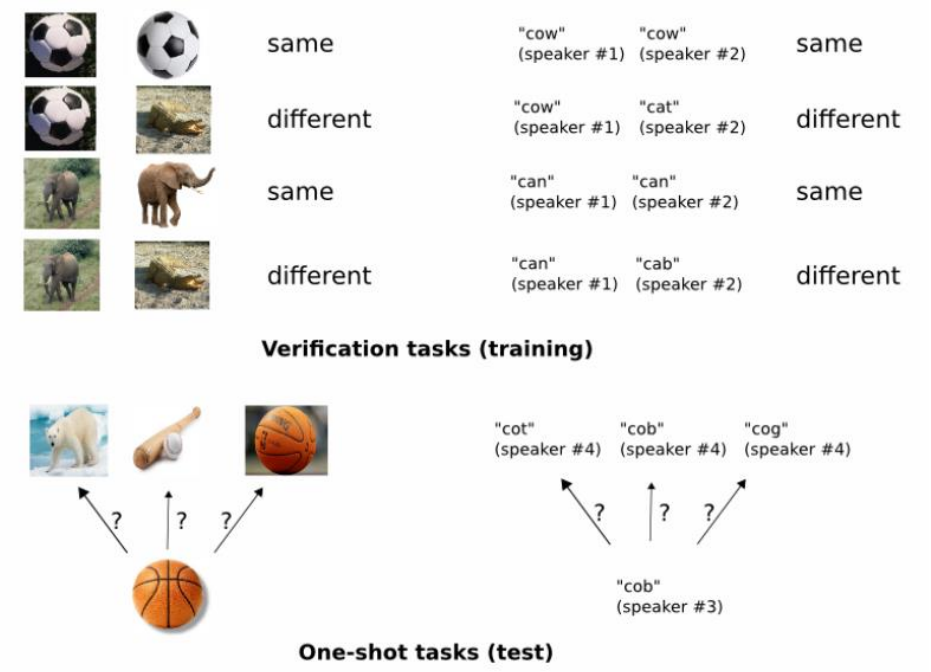
### 2.1.7 การเรียนรู้ด้วยข้อมูลจำนวนน้อย (Few-Shot Learning) และการเรียนรู้โดยปราศจากข้อมูลปฐมภูมิ (Zero-Shot Learning)

ในปัจจุบัน เพื่อสร้างแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่อง จำเป็นต้องใช้ชุดข้อมูลขนาดใหญ่เพื่อปรับค่าน้ำหนักและค่าความเอนเอียงของแบบจำลอง เพื่อให้การทำนายของแบบจำลองมีความแม่นยำและคล้ายคลึงกับผลลัพธ์จริงมากที่สุด การใช้ชุดข้อมูลขนาดใหญ่ช่วยลดค่าสูญเสีย (Loss) และเพิ่มความแม่นยำ (Accuracy) ของแบบจำลองได้ โดยมีความเชื่อมั่นว่า แบบจำลองที่ถูกฝึกด้วยชุดข้อมูลมากเพียงพอจะมีความสามารถในการทำนายผลที่แม่นยำมากขึ้น เนื่องจากมีข้อมูลมากขึ้นที่สามารถสอนแบบจำลองได้ดีขึ้น อย่างไรก็ตาม การฝึกแบบจำลองด้วยชุดข้อมูลขนาดใหญ่มีความต้องการต้นทุนที่สูง ไม่ว่าจะเป็นค่าใช้จ่ายในการรวบรวมข้อมูล ค่าใช้จ่ายในการทำความสะอาดข้อมูล และเวลาในการฝึกแบบจำลองที่เพิ่มขึ้น จึงต้องมองหาวิธีการลดต้นทุนในการพัฒนา ซึ่งการเรียนรู้ด้วยข้อมูลจำนวนน้อย (Few-shot Learning) และการเรียนรู้โดยปราศจากข้อมูลปฐมภูมิ (Zero-shot Learning) เป็นวิธีหนึ่งที่สามารถนำมาช่วยลดต้นทุนในการพัฒนาได้

จากการศึกษาได้พบว่ามนุษย์มีวิธีการเรียนรู้ที่ใกล้เคียงกับเครื่องมาก โดยมนุษย์สามารถใช้ตัวอย่างเพียงไม่กี่ตัวอย่างเท่านั้นในการจดจำและแยกแยะข้อมูล เช่น สามารถแยกแยะตัวเลขที่เกิดจากลายมือได้ด้วยการเรียนรู้จากไม่กี่ ภาพเท่านั้น แต่สำหรับเครื่องการเรียนรู้จำเป็นต้องใช้ภาพตัวอย่างจำนวนมากเพื่อให้ แบบจำลองสามารถเรียนรู้ได้อย่างมีประสิทธิภาพ แนวคิดนี้ถูกนำเสนอในแบบจำลองการจดจำด้วยภาพ (Image Recognition) ซึ่งได้รับความนิยมในงานวิจัย (Koch, Zemel, & Salakhutdinov, 2015) แบบจำลองในการจดจำด้วยภาพทำงานโดยการเรียนรู้และสกัดค่าสัญญาณจากภาพ และนำมาเปรียบเทียบเพื่อหาความคล้ายคลึงและความต่างของภาพในปริภูมิเวกเตอร์ จากนั้นแบบจำลองจะทำนายว่าภาพที่เป็นข้อมูลนำเข้าเหมือนหรือต่างกัน การทำนายนี้เป็นการทำนายแบบจำแนกข้อมูลสองประเภท (Binary Classification) เมื่อแบบจำลองเรียนรู้เสร็จสิ้น จะทดสอบโดยใช้คู่ของภาพที่ไม่เคยเห็นมาก่อนและให้แบบจำลองทำนายว่าภาพนั้นเหมือนหรือแตกต่างกัน แบบจำลองจะค้นหาค่าสัญญาณของทั้งสองภาพ และเปรียบเทียบค่าสัญญาณว่าตรงกันหรือไม่ หากมีความคล้ายคลึงหรือตรงกัน แบบจำลองจะทำนายว่าเป็นภาพที่ใกล้เคียง แต่

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ถ้าไม่มีความคล้ายคลึง แบบจำลองจะทำนายว่าเป็นภาพที่ไม่ใกล้เคียง ตัวอย่างการทำงานของแบบจำลองแสดงในรูปที่ 12




*รูปภาพ 12 แผนภาพแสดงการทำงานของแบบจำลอง (Koch, Zemel et al. 2015)*

เทคนิคการเรียนรู้ด้วยข้อมูลจำนวนน้อย (Few-shot Learning) และการเรียนรู้โดยปราศจากข้อมูลปฐมภูมิ (Zero-shot Learning) เป็นวิธีที่ช่วยลดต้นทุนในการพัฒนา โดยใช้การเรียนรู้จากข้อมูลที่มีหรือข้อมูลในโดเมน (Domain) ที่จำกัดเพื่อนำไปประยุกต์ใช้กับโดเมนอื่น ๆ ตัวอย่างเช่นการเรียนรู้จากชุดข้อมูลของกฎหมายละเมิด ทำให้สามารถทำนายผลการเลือกกฎหมายละเมิดได้ด้วยความแม่นยำที่ค่อนข้างสูง แบบจำลองที่เรียนรู้จากชุดข้อมูลเหล่านี้ยังสามารถนำไปใช้ในการเลือกกฎหมายแพ่งหรือพาณิชย์ที่เกี่ยวข้องอื่น ๆ ที่เป็นข้อมูลใหม่ที่แบบจำลองไม่เคยเรียนรู้ได้อย่างแม่นยำ โดยการเรียนรู้ในลักษณะนี้มีความคล้ายคลึงกับการเรียนรู้ของมนุษย์มาก โดยมนุษย์สามารถใช้ตัวอย่างน้อย ๆ ในการเรียนรู้และจดจำข้อมูลได้ เช่นการแยกตัวเลขที่เกิดจากลายมือ มนุษย์สามารถแยกได้ด้วยการเรียนรู้จำนวนภาพไม่มากนัก แต่สำหรับเทคนิคการเรียนรู้ของเครื่องจำเป็นต้องใช้ภาพตัวอย่างจำนวนมากเพื่อให้แบบจำลองเรียนรู้

ในกรณีของการเรียนรู้โดยปราศจากข้อมูลปฐมภูมิ แบบจำลองจะถูกฝึกด้วยข้อมูลนำเข้าที่เป็นประมวลกฎหมายแพ่งและพาณิชย์ที่เกี่ยวข้องกับคดีละเมิด และทดสอบกับกฎหมายแพ่งและพาณิชย์ตัวอื่นที่ไม่เกี่ยวข้องกับคดีละเมิดเลย แบบจำลองจะไม่เคยเห็น

ข้อมูลรับเข้าในลักษณะนี้มาก่อน จากนั้นแบบจำลองจะคำนวณความแม่นยำเพื่อประเมินผล

**2.1.8 การประเมินผลแบบจำลอง**

ในขั้นตอนการพัฒนาแบบจำลอง ข้อมูลที่ใช้จะถูกแบ่งออกเป็นชุดฝึกฝน (training set) ชุดพัฒนา (development set) และชุดทดสอบ (test set) โดยแต่ละชุดข้อมูลมีหน้าที่แตกต่างกันในกระบวนการพัฒนาและประเมินแบบจำลอง ดังนี้:

**ชุดฝึกฝน (Training set)** เป็นชุดข้อมูลที่ใช้ในขั้นตอนการเรียนรู้ของแบบจำลอง แบบจำลองจะใช้ชุดข้อมูลฝึกฝนเพื่อปรับค่าพารามิเตอร์และน้ำหนักของตัวเองเพื่อให้ได้ผลการทำนายที่ดีขึ้น ในการสร้างชุดฝึกฝน จะใช้ข้อมูลที่มีป้ายกำกับอย่างถูกต้องเพื่อให้แบบจำลองเรียนรู้และปรับตัวให้ถูกต้องตามข้อมูลที่ให้มา

**ชุดพัฒนา (Development set)** เป็นชุดข้อมูลที่ใช้ในขั้นตอนการพัฒนาและปรับปรุงแบบจำลอง แบบจำลองจะถูกทดสอบกับชุดพัฒนาเพื่อวัดประสิทธิภาพและปรับแก้ไขไฮเปอร์พารามิเตอร์เพื่อเพิ่มความแม่นยำในการทำนาย ชุดพัฒนาจะมีป้ายกำกับเช่นเดียวกับชุดฝึกฝนแต่จะไม่มีการใช้ในกระบวนการปรับแก้ไขแบบจำลอง

**ชุดทดสอบ (Test set)** เป็นชุดข้อมูลที่ใช้ในการประเมินผลลัพธ์ของแบบจำลองที่พัฒนาขึ้น แบบจำลองที่ได้จากขั้นตอนการฝึกฝนและพัฒนาจะถูกทดสอบกับชุดทดสอบเพื่อดูประสิทธิภาพและความสามารถในการทำนาย ชุดทดสอบควรเป็นข้อมูลที่แบบจำลองไม่เคยเห็นมาก่อนเพื่อให้ผลการทดสอบแสดงถึงประสิทธิภาพการทำนายที่แท้จริงของแบบจำลอง

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ในการประเมินแบบจำลองการจำแนกเอกสารที่มีการทำนายเป็น 2 รูปแบบ มักนิยมใช้เครื่องมือวัดประสิทธิภาพต่าง ๆ เพื่อวัดผลการทำนาย ที่นิยมนำมาใช้คือค่าความแม่นยำ (accuracy) ค่าความเที่ยงตรง (precision) ค่าความถูกต้อง (recall) และ f1-score เพื่อวัดประสิทธิภาพก ารทำนายโดยรวมของแบบจำลอง โดยสามารถคำนวณแต่ละตัวดังนี้:

*2.1.8.1 Accuracy (ความแม่นยำ)*

คืออัตราส่วนระหว่างจำนวนครั้งที่แบบจำลองทำนายตรงกับป้ายกำกับที่ถูกต้องต่อจำนวนครั้งทั้งหมดที่แบบจำลองทำนาย ซึ่งเป็นตัววัดที่แสดงถึงความถูกต้องของแบบจำลองในการทำนายทั้งหมด

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

*2.1.8.2 Precision (ความเที่ยงตรง)*

คืออัตราส่วนระหว่างจำนวนครั้งที่แบบจำลองทำนายผลเป็นป้ายกำกับหนึ่งที่ถูกต้องต่อจำนวนครั้งที่แบบจำลองทำนายผลเป็นป้ายกำกับหนึ่งทั้งหมด ซึ่งจะแสดงถึงความเที่ยงตรงในการทำนายป้ายกำกับแต่ละประเภทว่าแบบจำลองมีความถูกต้องตามข้อมูลเช่นนั้นหรือไม่ โดยที่แบบจำลองไม่ได้ทำนายได้ดีโดยเข้าใจเพียงแค่การสุ่ม

$$\frac{TP}{TP + FP} \tag{8}$$

*2.1.8.3 Recall (ความถูกต้อง)*

คืออัตราส่วนระหว่างจำนวนครั้งที่แบบจำลองทำนายผลเป็นป้ายกำกับหนึ่งที่ถูกต้องต่อจำนวนครั้งที่แบบจำลองทำนายผลป้ายกำกับทุกรูปแบบที่ถูกต้อง ซึ่งจะแสดงถึงความสามารถของแบบจำลองในการทำนายได้ถูกต้องในแต่ละป้ายกำกับโดยไม่มีความเอนเอียงไปที่ป้ายกำกับประเภทใดๆ

$$\frac{TP}{TP + FN} \tag{9}$$

*2.1.8.4 F1-score (คะแนน F1)*

คือค่าเฉลี่ยฮาร์โมนิกของค่า precision และค่า recall โดยคำนวณตามสมการดังนี้:

$$F1 = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall}\right) \tag{10}$$

โดยการใช้เครื่องมือวัดเหล่านี้จะช่วยวัดประสิทธิภาพของแบบจำลองในการทำนายและเปรียบเทียบกับแบบจำลองอื่น ๆ ที่ถูกทดสอบในการทดลอง ผลลัพธ์เหล่านี้จะช่วยให้ทราบถึงประสิทธิภาพการทำนายโดยรวมของแบบจำลองที่พัฒนาขึ้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การค้นคืนสารสนเทศ (Information Retrieval) เป็นกระบวนการค้นหาและแสดงผลลัพธ์ที่เกี่ยวข้องกับคำค้นหาหรือข้อมูลที่ผู้ใช้ป้อนเข้ามา ในการวัดประสิทธิภาพของระบบการค้นคืนสารสนเทศ ประเด็นสำคัญคือการประเมินความสามารถในการแสดงผลลัพธ์ที่เกี่ยวข้องกับคำค้นหาหรือข้อมูลที่ผู้ใช้ต้องการหา ในทางปฏิบัติ มีการใช้มาตรฐานการประเมินผลที่นิยมนำมาใช้ในการทดสอบแบบจำลองการค้นคืนสารสนเทศ ซึ่งเป็น f1-score@k โดยที่ k คือจำนวนผลลัพธ์ที่จะพิจารณา f1-score@k เป็นการวัดประสิทธิภาพการทำนายโดยใช้เกณฑ์เฉลี่ย ฮาร์โมนิก ของค่า precision และค่า recall โดยการคำนวณมีลักษณะเดียว กับที่กล่าวไปแล้ว โดยในการประเมินผลด้วย f1-score@k จะนำผลการทำนายที่มีความน่าจะเป็นสูงสุด k อันดับในแต่ละคำค้นหาหรือเอกสารมาใช้ในการ

คำนวณ ตัวอย่างเช่น หากต้องการประเมินประสิทธิภาพในการทำนายมาตรา 55 และมาตรา 242 จากแบบจำลองตัวอย่างในตารางที่ 2.1 โดยใช้ f1-score@2 จะคำนวณค่า f1-score จากผลการทำนายมาตราสูงสุด 2 อันดับในแต่ละเอกสารทางกฎหมายได้ดังนี้

หากค่า precision และค่า recall สำหรับการทำนายมาตรา 55 ในเอกสารที่ 1 คือ 0.8 และ 0.6 ตามลำดับ และสำหรับการทำนายมาตรา 242 ในเอกสารที่ 1 คือ 0.7 และ 0.5 ตามลำดับ โดยมีเอกสารทั้งหมด 5 เอกสาร ผลการคำนวณ f1-score@2 จะเป็นดังนี้:

$$\text{f1-score@2} = 2 * (0.8 * 0.6 + 0.7 * 0.5) / (0.8 + 0.6 + 0.7 + 0.5) = 0.64$$

การคำนวณนี้จะช่วยให้ประเมินความสามารถในการทำนายและค้นหาข้อมูลที่เกี่ยวข้องในแต่ละคำค้นหาหรือเอกสาร และจะช่วยในการเปรียบเทียบความสามารถของแบบจำลองการค้นคืนสารสนเทศที่พัฒนาขึ้น

## 2.2 งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยเกี่ยวกับการวิเคราะห์คำพิพากษาด้วยการประมวลผลภาษาธรรมชาติ (Natural Language Processing) นั้นมีการศึกษาและวิจัยที่ เกี่ยวข้องกันทั้งในต่างประเทศและในประเทศไทย ซึ่งสามารถแบ่งงานวิจัยดังกล่าว เป็นสองกลุ่มหลัก คือ งานวิจัยที่เกี่ยวกับการวิเคราะห์คำพิพากษาของศาลในต่าง ประเทศและงานวิจัยเกี่ยวกับการวิเคราะห์คำพิพากษาของศาลในประเทศ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

### 2.2.1 งานวิจัยเกี่ยวกับการวิเคราะห์คำพิพากษาของศาลในต่างประเทศ

งานศึกษาเกี่ยวกับคำพิพากษาของศาลต่างประเทศเป็นหนึ่งในงานวิจัยที่น่าสนใจ โดยเฉพาะอย่างยิ่งเมื่อพูดถึงประเทศสหรัฐอเมริกาซึ่งเป็น ประเทศที่มีการศึกษา และ การวิเคราะห์คำพิพากษาจำนวนมาก ตัวอย่างที่น่าสนใจ เป็นงานศึกษาของ (Katz, Bommarito et al. 2017) ที่ได้ทำการศึกษาเกี่ยวกับพฤติกรรม การตัดสินของผู้พิพากษาในศาล สูงสุด (Supreme Court) ซึ่งเน้นการวิเคราะห์การยืนยันหรือกลับคำพิพากษา รวมถึงการวิเคราะห์คำตัดสินของผู้พิพากษาแต่ละคน โดยใช้ชุดข้อมูลจากฐานข้อมูล Supreme Court Database (SCDB) ซึ่งประกอบไปด้วยข้อมูลเกี่ยวกับศาลและ คำพิพากษาตั้งแต่ปี ค.ศ. 1816 ถึง ค.ศ. 2015 รวมทั้งผลการโหวต ของผู้พิพากษามากถึง 240,000 ครั้ง และผลคำพิพากษาของคดีทั้งหมด 28,000 คดี จากการทดสอบแบบจำลองที่ดำเนินการ ผลการทดสอบแสดงให้เห็นถึงความแม่นยำสูงถึงร้อยละ 70.2 ในการทำนายคำพิพากษารายคดีและความแม่นยำสูงถึงร้อย

ละ 71.9 ในการทำนายคำพิพากษารายคน งานวิจัยยังสรุปว่าการทำนายคำพิพากษาจะมีผลประโยชน์อย่างมากเนื่องจากคำพิพากษาส่งผลต่อกลุ่มบุคคลในสังคมจำนวนมากทั้งในภาครัฐและภาคเอกชน นอกจากนี้ ยังมีประเทศอื่นๆ ที่กำลังเริ่มศึกษาและนำปัญหาดังกล่าวมาวิเคราะห์คำพิพากษา เช่น ประเทศฝรั่งเศสที่มีการวิเคราะห์คำพิพากษาของศาลสูงสุดในทิศทางเดียวกัน

นอกจากนี้มีการนำปัญญาประดิษฐ์ (AI) มาใช้กับกฎหมายเฉพาะด้านอีกด้วย คณะทำงานจากมหาวิทยาลัยโทรอนโทได้ริเริ่มโครงการวิจัยที่เรียกว่า "Blue J Legal" (Alarie and Gardhouse 2021) ซึ่งมุ่งเน้นการใช้ปัญญาประดิษฐ์ในการวิเคราะห์กฎหมายภาษีของประเทศแคนนาดา ซึ่งเป็นกฎหมายที่มีความไม่ชัดเจนในตัวบทหรือมาตรฐานทางกฎหมายหลายแห่ง อย่างไรก็ตาม โครงการนี้ได้ทำการวิเคราะห์ตัวอย่างข้อกฎหมายที่ท้าทาย เช่น คำถามเกี่ยวกับลูกค้าซึ่งเป็นลูกจ้างหรือผู้รับจ้างอิสระ (independent contractors) ซึ่งเป็นคำถามสำคัญในกฎหมายกำหนดเรื่องภาษีเงินได้ ผู้วิจัยได้ใช้ปัญญาประดิษฐ์ในการค้นหาความสัมพันธ์และน้ำหนักของแต่ละปัจจัยในกฎหมายเพื่อจัดประเภทบุคคลตามกฎหมายภาษี ผลลัพธ์ที่ได้มีความแม่นยำถึงร้อยละ 90 นอกจากนี้ ระยะเวลาที่คอมพิวเตอร์ใช้ในการวิจารณาก็สั้นกว่าระยะเวลาที่ทนายความใช้ในการพิจารณาอย่างมาก

อีกหนึ่งตัวอย่างของงานศึกษาทางกฎหมายเฉพาะด้านคือกฎหมายที่เกี่ยวกับการลี้ภัย (Asylum court decision) (Dunn, Sagun et al. 2017) โดยมีจุดประสงค์ในการพัฒนาระบบปัญหาประดิษฐ์เพื่อช่วยผู้ที่กำลังจะยื่นขอต่อศาลให้พิจารณาสถานะ เพื่อให้บุคคลเหล่านั้นทราบถึงความเป็นไปได้ในการขอสถานะของตนเองรวมถึงให้ทราบถึงปัจจัยที่มีผลกระทบต่อความสำเร็จของการยื่นขอด้วยความแม่นยำของระบบ ในงานศึกษานี้สรุปได้ว่าระบบปัญหาประดิษฐ์มีความแม่นยำสูงถึงร้อยละ 80 นอกจากนี้งานศึกษายังค้นพบปัจจัยที่มีผลกระทบสูงและสามารถระบุสัดส่วนของผู้พิพากษาที่มีความแน่นอนในวิธีการตัดสินคดีได้อีกด้วย

ศาลระหว่างประเทศก็เป็นอีกหนึ่งประเภทที่ได้รับการวิเคราะห์โดยการปัญญาประดิษฐ์เช่นกัน ในปี ค.ศ. 2016 (Aletras, Tsarapatsanis et al. 2016) ได้มีงานศึกษาที่เกี่ยวกับการพัฒนาแบบจำลองของผลคำพิพากษาของศาลสิทธิมนุษยชนแห่งยุโรป (The European Court of Human Rights: ECHR) โดยนำข้อมูลจริงและคำพิพากษาภายใต้อนุสัญญายุโรปว่าด้วยสิทธิมนุษยชน (The European Convention of Human Rights) เป็นข้อมูลในการสร้างแบบจำลองที่ทำนายว่าข้อเท็จจริงเป็นการละเมิดตามอนุสัญญาหรือไม่ ผลการทดสอบแบบจำลองพบว่ามีความแม่นยำสูงในการทำนายผลการตัดสินในมาตราที่

ต้องการ โดยร้อยละ 78 สำหรับมาตรา 3 (Prohibition of torture) ร้อยละ 84 สำหรับมาตรา 6 (Right to a fair trial) และร้อยละ 79 สำหรับมาตรา 8 (Private and family life) นอกจากนี้ยังได้สังเกตเกี่ยวกับปัจจัยที่สำคัญที่สุดในการพิจารณาคือข้อเท็จจริงของคดี (Factual background) ซึ่งอยู่ในคำพิพากษาของศาลสิทธิมนุษยชนยุโรป

ในปี ค.ศ. 2019 (Chalkidis, Androutsopoulos et al. 2019) มีการศึกษาและสร้างแบบจำลองเพื่อพยากรณ์คำพิพากษาของศาลสิทธิมนุษยชนแห่งยุโรป (The European Courts of Human Rights: ECHR) โดยคณะผู้วิจัยรวบรวมชุดข้อมูลจากฐานข้อมูลสาธารณะนี้ที่ประกอบด้วยคดีทั้งหมดประมาณ 11,500 คดี และสร้างแบบจำลองในลักษณะที่แตกต่างกันออกไป 3 แบบ ได้แก่

1. การจำแนกประเภทการละเมิดแบบไบนารี (binary violation classification) โดยใช้ข้อเท็จจริงของคดี แบบจำลองจะแบ่งประเภทการละเมิดเป็นเชิงบวกถ้ามีการละเมิดกฎหรือมาตราใด ๆ ของอนุสัญญาสิทธิมนุษยชน และเป็นเชิงลบถ้าไม่มีการละเมิด ผลการประเมินของแบบจำลองนี้ได้ค่า F1 อยู่ที่ร้อยละ 80.5
2. การจำแนกประเภทการละเมิดแบบหลายป้ายกำกับ (multi-label classification) โดยใช้ข้อเท็จจริงของคดี แบบจำลองจะเลือกว่าคดีละเมิดกฎหรือมาตราใด ๆ ของอนุสัญญาสิทธิมนุษยชน จากทั้งหมด 66 ข้อ และจะไม่เลือกกฎใด ๆ หากไม่ได้ละเมิด ผลการประเมินของแบบจำลองนี้ได้ค่า F1 อยู่ที่ร้อยละ 60.8
3. การทำนายความสำคัญของคดี (case importance prediction) คณะผู้วิจัยทำนายความสำคัญของคดีโดยให้คะแนนตั้งแต่ 1 (มีความสำคัญ) ถึง 4 (ไม่มีความสำคัญ) โดยใช้การถดถอย (regression) โดย ECHR ได้สร้างชุดข้อมูลทดสอบที่ประกอบด้วย 1096 เอกสารที่ได้คะแนน 1, 904 เอกสารที่ได้คะแนน 2,982 เอกสารที่ได้คะแนน 3, และ 6,496 เอกสารที่ได้คะแนน 4 ซึ่งแสดงให้เห็นว่าประมาณ 10% เป็นคดีที่สำคัญ ในขณะที่ส่วนใหญ่ (83%) มีความสำคัญน้อยลงตามลำดับ ผลการประเมินของแบบจำลองนี้ได้ค่าเฉลี่ยของความผิดพลาดอยู่ที่ประมาณ 0.437

**2.2.2 งานวิจัยเกี่ยวกับคำพิพากษาของศาลในประเทศไทย**

ในปีค.ศ. 2018 (Kowsrihawat, Vateekul et al. 2018) มีการศึกษาคำพิพากษาที่เกี่ยวข้องกับกฎหมายอาญา โดยรวบรวมจากเว็บไซต์ระบบสืบค้นคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา (ศูนย์เทคโนโลยีสารสนเทศและการสื่อสารในศาลฎีกา, 2017) ซึ่งประกอบไปด้วยคำพิพากษาและข้อกฎหมายจำนวนคดีอาญาที่ใช้ในการศึกษาทั้งหมด 1,000

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ฉบับ ระหว่างปีค.ศ. 1958 ถึงปีค.ศ. 2016 คณะผู้วิจัยกำหนดขอบเขตของการศึกษาในการสร้างชุดข้อมูลทดสอบ Thai Supreme court cases (TSCC) และสร้างแบบจำลองเพื่อทำนายผลการตัดสินคดีอาญาโดยใช้การเทคนิคการเรียนรู้เชิงลึก โดย TSCC แบ่งประเภทความผิดอาญาเป็น 3 หมวดหลัก ได้แก่ ความผิดเกี่ยวกับชีวิตและร่างกาย (11 มาตรา) ความผิดเกี่ยวกับชื่อเสียง (3 มาตรา) และความผิดเกี่ยวกับทรัพย์สิน (15 มาตรา) แบบจำลองที่พัฒนามีความแม่นยำถึง 64.63% นอกจากนี้ คณะผู้วิจัยได้ทำการทบทวนงานวิจัยในการวิเคราะห์คำพิพากษาของศาลในประเทศไทย เช่น ในปีค.ศ. 2008 (Thammaboosadee and Silparcha 2008) ได้ทำการวิเคราะห์และพัฒนาระบบการให้เหตุผลในคำตัดสินคดีอาญา โดยใช้เทคนิคเหมืองข้อมูลจากชุดข้อมูล Thai Court XML (TCXML) ซึ่งเป็นการรวบรวมและสกัดองค์ประกอบความผิดจากข้อเท็จจริง โดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision tree) ในปี 2014 (Thammaboosadee, Watanapa et al. 2014) ได้พัฒนาระบบจำแนกประเภทสองขั้นตอนเพื่อระบุฐานความผิดและการลงโทษในกฎหมายอาญาจาก TCXML โดยแบ่งออกเป็นสองส่วน คือ การทำนายคุณลักษณะของคดี ซึ่งคุณลักษณะของคดีนั้นที่อยู่ด้วยกัน 25 คุณลักษณะ เช่น เจตนาของการกระทำ การใตร่ตรองไว้ก่อน เป็นต้น จาก 25 คุณลักษณะทางคณะผู้วิจัยใช้ PCA เพื่อลดขนาดข้อมูลเข้าสู่ระบบ 13 คุณลักษณะ และนำข้อมูลเข้าไปเรียนรู้โดยใช้โครงข่ายประสาทเทียม ส่วนในส่วนที่สอง ผู้วิจัยจะทำนายคุณลักษณะทางกฎหมายจาก 25 คุณลักษณะก่อนหน้าเพื่อให้ได้ 9 คุณลักษณะทางกฎหมาย เช่น เจตนาในการกระทำความผิด การบันดานโทสะ เป็นต้น โดยใช้เทคนิคการเรียนรู้ SVM-RFE ที่เป็น SVM แบบมีความสามารถในการจัดลำดับข้อมูลตัวแทน (Feature ranking)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

# บทที่ 3
# แนวคิดและวิธีการวิจัย

## 3.1 การสร้างคลังข้อมูลคำพิพากษา

ในปัจจุบันประเทศไทยมีการเผยแพร่ข้อมูลคำพิพากษาอย่างแพร่หลาย แต่มีลักษณะที่แตกต่างกันออกไปอย่างชัดเจน เช่น รูปแบบแรก คือ เอกสารรูปแบบฉบับพิมพ์ (hardcopy) ซึ่งต้องทำการขอสำเนาจากศาลในกรณีที่ต้องการข้อมูลประเภทนี้ ส่วนรูปแบบที่สอง คือ ข้อมูลอิเล็กทรอนิกส์ที่ยากต่อการวิเคราะห์ด้วยคอมพิวเตอร์ เช่น ไฟล์รูปแบบ PDF หรือแม้แต่การเผยแพร่ผ่านเว็บไซต์โดยใช้รูปแบบ HTML ซึ่งยากต่อการอ่านโดยเครื่องคอมพิวเตอร์ ดังนั้น คณะผู้วิจัยเห็นว่าการสร้างคลังข้อมูลคำพิพากษาเป็นขั้นตอนสำคัญสำหรับงานวิจัยนี้ เพื่อให้เป็นที่เก็บข้อมูลที่เข้าถึงง่าย นอกจากนี้ยังมีแนวทางในการสร้างคลังข้อมูลที่อำนวยความสะดวกในการวิเคราะห์คำพิพากษา โดยการนำเอาข้อมูลเช่น เลขที่คำพิพากษา ปีที่ตัดสิน หรือส่วนต่างๆของเนื้อความในคำพิพากษามาใช้ เพื่อให้เป็นแนวทางที่สามารถใช้ในการสร้างคลังข้อมูลที่ใหญ่ขึ้น เพื่อให้ผู้สนใจและหน่วยงานของรัฐสามารถนำข้อมูลไปวิเคราะห์เพื่อเกิดประโยชน์ต่อไปได้

## 3.2 ขั้นตอนการรวบรวมคำพิพากษา

ในกระบวนการรวบรวมคำพิพากษา คณะผู้วิจัยเลือกใช้เทคนิค Web Scraping เพื่อดึงข้อมูลจากเว็บไซต์ระบบสืบค้นคำพิพากษา คำสั่งคำร้อง และคำวินิจฉัยศาลฎีกาซึ่งสามารถเข้าถึงได้จาก http://deka.supremecourt.or.th/ โดยใช้เครื่องมือ PhantomJs (https://phantomjs.org/) เพื่อค้นหาและคัดลอกข้อมูลต่างๆ และบันทึกในรูปแบบของภาษามาร์กอัปเอ็มแอล (HTML) ซึ่งสามารถนำมาประมวลผลต่อไปได้ เมื่อได้รับข้อมูลทั้งหมด คณะผู้วิจัยได้สร้างเครื่องมือโดยใช้ภาษา Python และ BeautifulSoup (https://www.crummy.com/software/BeautifulSoup/) เพื่อแปลงข้อมูลจากรูปแบบ HTML เป็นโครงสร้างที่เหมาะสมสำหรับการค้นหา การจัดหมวดหมู่ และการประมวลผลอื่นๆ นักวิจัยทางด้านกฎหมายจากทีมได้รับหน้าที่ตรวจสอบความถูกต้องของข้อมูล และโครงสร้างข้อมูลสรุปถูกจัดเก็บในตารางที่ 1

*ตาราง 1 แสดงโครงสร้างของการจัดเก็บข้อมูลคำพากษา*

| **โครงสร้าง** | **คำอธิบาย** |
|---|---|
| case_no | เลขที่คำพิพากษา |
| case_year | ปีที่ตัดสิน |

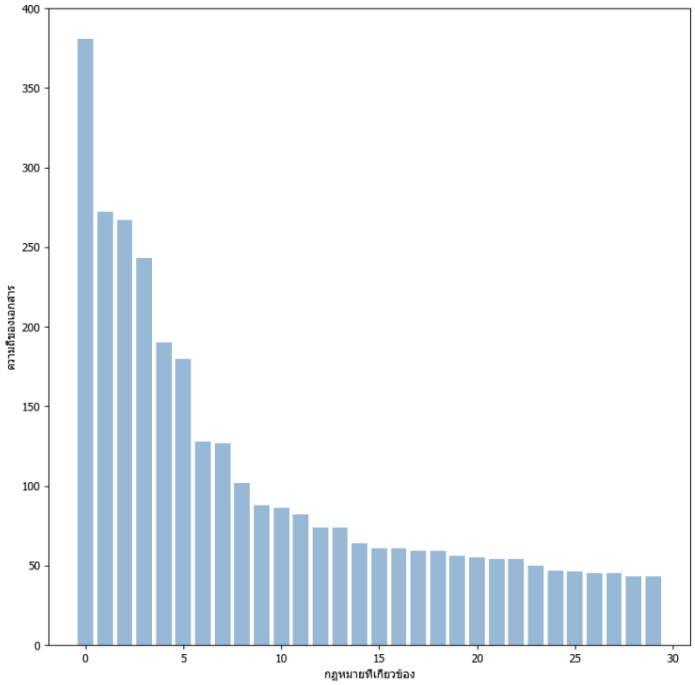
| | |
|---|---|
| sort_text | ย่อสั้น |
| long_text | ย่อยาว |
| has_long | Flag ที่จะบอกว่ามีหรือไม่มีย่อยาวเพื่อความรวดเร็วในการกรองข้อมูล |
| article | กฎหมายที่เกี่ยวข้อง เช่น ประมวลกฎหมายแพ่งและพาณิชย์ |
| short_article | ตัวย่อของกฎหมายที่เกี่ยวข้อง เช่น ป.พ.พ. |
| section | มาตรา ของกฎหมายที่เกี่ยวข้อง เช่น มาตรา 420 |
| subsection | อนุมาตรา ของกฎหมายที่เกี่ยวข้อง เช่น (2) |
| paragraph | วรรค ของกฎหมายที่เกี่ยวข้อง เช่น วรรคสอง |

ภายหลังจากการประมวลผลและตรวจสอบแล้ว จะได้คลังข้อมูลคำพิพากษาที่มีขนาด 80,654 คำพิพากษา ซึ่งมี 206,530,463 ตัวอักษร ตั้งแต่ปี 2471 ถึง 2559

**3.3 การคัดแยกคำพิพากษาที่เกี่ยวข้องกับคดีละเมิด**

เพื่อให้สามารถสร้างเครื่องจำแนกกฎหมายที่เกี่ยวข้องกับคดีละเมิดได้ คณะผู้วิจัยได้กำหนดโจทย์ในการใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) โดยต้องการข้อมูลที่เป็นข้อเท็จจริงและตัวบทของกฎหมายเพื่อนำมาป้อนให้แก่ระบบ (Input) และต้องการให้ระบบทำนายกฎหมายที่เกี่ยวข้อง (Label) โดยอ้างอิงถึงบทบัญญัติว่าด้วยหลักทั่วไปและคำนิยามเกี่ยวกับละเมิดตามประมวลกฎหมายแพ่งและพาณิชย์มาตรา 420 ซึ่งเป็นหลักความผิดที่มักจะนำมาใช้ในคดีละเมิดอยู่ส่วนใหญ่ ซึ่งคณะผู้วิจัยได้ดำเนินการสำรวจสถิติเกี่ยวกับมาตรากฎหมายที่เกี่ยวข้องกับคดีละเมิด ซึ่งจากการคัดแยกจะได้คำพิพากษาทั้งสิ้น 4,448 คำพิพากษา จากคำพิพากษาศาลฎีกาในคดีละเมิดที่มีอยู่ในฐานข้อมูลก่อนหน้านี้ โดยเฉพาะเนื้อหาของมาตรา 420 ซึ่งเป็นมาตราหลักที่ใช้ในกรณีคดีละเมิด มาตรานี้เป็นการระบุหลักความผิดที่สำคัญและมีการอ้างอิงกับมาตราอื่น ๆ ที่เกี่ยวข้อง ผลการสำรวจสถิติพบว่ามาตรากฎหมายที่เกี่ยวข้องแสดงพฤติกรรมการกระจายตัวที่มีลักษณะเป็น Long tail ซึ่งแสดงให้เห็นถึงการกระจายตัวที่แตกต่างกันอย่างมาก รูปภาพด้านล่างแสดงการกระจายตัวดังกล่าวในรูปแบบกราฟ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY




*รูปภาพ 13 สถิติมาตรากฎหมายที่เกี่ยวข้องกับคดีละเมิดจากวิเคราะห์คลังข้อมูล*

*ตาราง 2 แสดงมาตรา 10 อันดับที่พบในฐานข้อมูลเรียงจากความถี่มากไปน้อย*

| อันดับ | ความถี่ | ข้อกฎหมาย |
|---|---|---|
| 1 | 35 | ประมวลกฎหมายวิธีพิจารณาความแพ่ง มาตรา 55 การโต้แย้งสิทธิทางศาล |
| 2 | 30 | ประมวลกฎหมายแพ่งและพาณิชย์ มาตรา 438 ค่าสินไหมทดแทนให้ศาลวินิจฉัยตามควรแก่พฤติการณ์และความร้ายแรงแห่งละเมิด |
| 3 | 20 | ประมวลกฎหมายวิธีพิจารณาความแพ่ง มาตรา 142 ห้ามตัดสินนอกคำฟ้อง |
| 4 | 17 | ประมวลกฎหมายวิธีพิจารณาความแพ่ง มาตรา 172 การฟ้องต้องชัดแจ้ง |
| 5 | 16 | ประมวลกฎหมายแพ่งและพาณิชย์ มาตรา 448 การนับอายุความในคดีละเมิด |
| 6 | 13 | ประมวลกฎหมายแพ่งและพาณิชย์ มาตรา 425 ความรับผิดของนายจ้าง |
| 7 | 13 | ประมวลกฎหมายแพ่งและพาณิชย์ มาตรา 1336 สิทธิใช้สอยและจำหน่ายทรัพย์สิน |
| 8 | 11 | ประมวลกฎหมายแพ่งและพาณิชย์ มาตรา 18 สิทธิในชื่อ |
| 9 | 11 | ประมวลกฎหมายแพ่งและพาณิชย์ มาตรา 880 การรับช่วงสิทธิสัญญาประกันภัย |

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

| | | |
|---|---|---|
| 10 | 10 | ประมวลกฎหมายแพ่งและพาณิชย์ มาตรา 850 ประนีประนอมยอมความ |

เมื่อทำการรวบรวมข้อมูลมาได้แล้วทางคณะผู้วิจัยได้แบ่งข้อมูลนำเข้าออกเป็นสองส่วนจากคำพิพากษา และยังมีข้อมูลเสริมเพิ่มเติมจากตัวบทบัญญัติของกฎหมายดังนี้

**3.3.1 ข้อมูลนำเข้าในส่วนของคำฟ้อง**

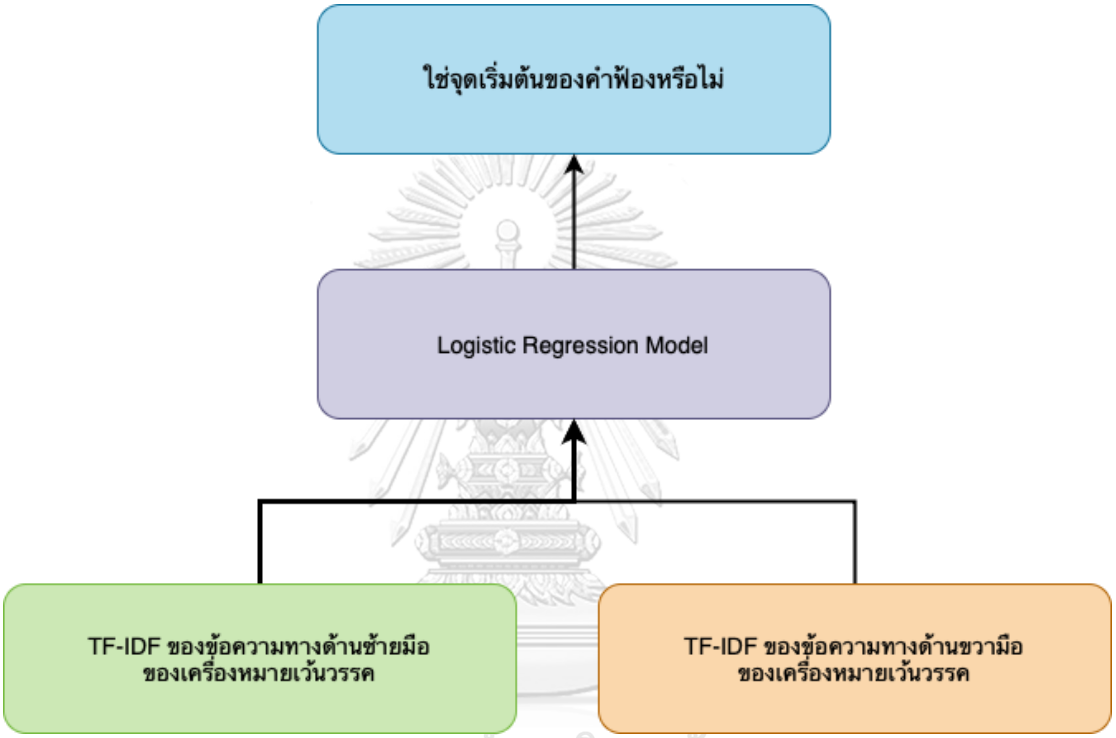
เนื่องจากคำพิพากษาที่ถูกเผยแพร่ออกมานั้นได้ถูกเขียนขึ้นจากเจ้าหน้าที่ในชั้นศาลและข้อมูลจำพวกคำฟ้องคำให้การจะถูกเขียนรวมกันเป็นเนื้อหาฉบับเดียว อาจจะมีหลายย่อหน้าหรือมีหลายประโยคแล้วแต่ความยาวในการตัดสินคดีความนั้นๆ เนื่องจากคณะผู้วิจัยมีผู้เชี่ยวชาญทางด้านกฎหมาย จึงได้ทำการสุ่มอ่านข้อความเพื่อพยายามหาข้อบ่งชี้ถึงคำฟ้องว่าอยู่บริเวณไหน เมื่อได้พยายามสุ่มเอกสารขึ้นมาอ่านแล้วก็ค้นพบว่ามีความเป็นไปได้หลากหลายมาก แต่ทั้งหมดจะอยู่ในย่อหน้าแรกของคำพิพากษา คณะผู้วิจัยทำการสร้างแบบจำลอง (Model) อย่างง่ายเพื่อใช้แยกคำฟ้องออกจาก โดยหลักการทำงานของแบบจำลองจะเป็นการตอบคำถามว่าใช่หรือไม่หรือที่เรียกกันว่าการจำแนกไบนารี่ (Binary Classification) โดยในแต่ละแบบจำลองนั้นจะประกอบด้วยข้อมูลนำเข้า (input) สองส่วน คือข้อมูลจากด้านซ้าย (ตัวอักษรสีแดงจากภาพที่ 14) ของเว้นวรรค และข้อมูลจากด้านขวาของเว้นวรรค (ตัวอักษรสีน้ำเงินจากภาพที่ 14) จากนั้นแบบจำลองจะทำการคำนวณว่าข้อความทางด้านขวาของเครื่องหมายเว้นวรรค (ระหว่างตัวอักษรสีแดงและสีน้ำเงินจากภาพที่ 14) นี้คือส่วนของคำฟ้องใช่หรือไม่

โจทก์ฟ้องและแก้ไขคำฟ้องว่า เมื่อวันที่ 25 กุมภาพันธ์ 2536จำเลยที่ 1 ซึ่งเป็นลูกจ้างขับรถยนต์บรรทุกหมายเลขทะเบียน 88-5610กรุงเทพมหานครไปในทางการที่จ้างของจำเลยที่ 2 ผู้เป็นนายจ้างแข่งกับรถยนต์บรรทุกหมายเลขทะเบียน 88-6345 กรุงเทพมหานคร บนถนนพหลโยธินจากกรุงเทพมหานครมุ่งหน้าไปทางจังหวัดสระบุรี ด้วยความประมาทใช้ความเร็วเกินกว่าอัตราที่กฎหมายกำหนด เป็นเหตุให้ชนท้ายรถยนต์บรรทุกคันดังกล่าวที่แล่นมาหยุดต่อท้ายรถยนต์บรรทุกหมายเลขทะเบียน70-0307 กรุงเทพมหานคร ของโจทก์ที่ติดการจราจรอยู่จนกระเด็นมาชนท้ายรถยนต์ของโจทก์ และรถยนต์ของโจทก์ไถลไปชนรถยนต์บรรทุกหมายเลขทะเบียน

*รูปภาพ 14 แสดงตัวอย่างของข้อมูลนำเข้าจำนวนหนึ่งเอกสาร*

เพื่อการทำนายว่าจุดใดเป็นจุดเริ่มต้นของคำพิพากษาหรือไม่ ในงานวิจัยนี้จะใช้วิธีการถดถอยลอจิสติก "Logistic regression" ดังภาพที่ 15 ซึ่งเป็นวิธีการที่พยายามทำนายว่าจุดใดเป็นจุดเริ่มต้นของคำพิพากษาโดยการสอนคอมพิวเตอร์ให้เรียนรู้จากคำที่อยู่ในบริเวณนั้นๆ โดยใช้ชุดข้อมูลที่เตรียมไว้ล่วงหน้าโดยมนุษย์ มีจำนวน 400 ชุด และจากนั้นคอมพิวเตอร์จะเรียนรู้ว่าลักษณะหรือรูปแบบ (pattern) ของชุดตัวเลขที่อยู่แวดล้อม

จุดเริ่มต้นหรือจุดสิ้นสุดของคำที่ถูกเรียนรู้นั้นมีอย่างไร ซึ่งจะช่วยในการระบุจุดเริ่มต้นหรือจุดสิ้นสุดของข้อเท็จจริงเมื่อได้รับข้อมูลคำพิพากษาใหม่ ในกรณีที่ชุดตัวเลขของเนื้อหาในบริเวณนั้นๆ มีความใกล้เคียงกันกับชุดตัวเลขของจุดเริ่มต้นหรือจุดสิ้นสุดของข้อเท็จจริง แบบจำลอง "Logistic Regression" จะถูกฝึกฝนด้วยคำพิพากษาโดยมีการเรียนรู้จนกระทั่งค่าความผิดพลาด (Error Rate) เปลี่ยนแปลงโดยไม่มีความสำคัญมีการเปลี่ยนแปลงที่ไม่มีนัยสำคัญ




*รูปภาพ 15 แสดงโครงสร้างของแบบจำลองในการระบุตำแหน่งเริ่มต้นของคำให้การ*

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

หลังจากการฝึกฝนแบบจำลองโดยใช้สมการถดถอยลอจิสติกและนำผลการทำนายจุดเริ่มต้นของคำฟ้องมาเปรียบเทียบกับตำแหน่งที่ถูกต้องของจุดเริ่มต้นของคำฟ้อง แบบจำลองได้ให้ค่าความแม่นยำ (Accuracy) เท่ากับ 0.996 ซึ่งหมายความว่าจาก 100 คำพิพากษา แบบจำลองสามารถทำนายได้ถูกต้องในเรื่องของจุดเริ่มต้นของข้อเท็จจริงถึง 99.6 คำพิพากษา นอกจากนี้ยังมีเมตริก (Metric) อื่น ๆ ที่แสดงผลดังนี้:

**Positive Precision:** 0.88 หมายความว่า ในข้อมูลที่แบบจำลองทำนายว่าเป็นจุดเริ่มต้นของคำฟ้องเป็นจริง มีความแม่นยำที่เท่ากับ 0.88

**Positive Recall:** 0.88 หมายความว่า แบบจำลองสามารถตรวจจับคำฟ้องที่เป็นจุดเริ่มต้นได้ในร้อยละ 88

**Positive F1:** 0.880 หมายความว่าค่าเฉลี่ยระหว่างความแม่นยำและความสามารถในการตรวจจับของแบบจำลองในการทำนายคำฟ้องเป็นจุดเริ่มต้นเท่ากับ 0.880

**Macro-averaged F1:** 0.94 หมายความว่าค่าเฉลี่ยระหว่างความแม่นยำและความสามารถในการตรวจจับของแบบจำลองทั้งหมดเท่ากับ 0.94

### 3.3.2 ข้อมูลนำเข้าในส่วนของคำให้การ

นอกเหนือจากคำฟ้องแล้ว คำให้การสามารถนำมาใช้เป็นข้อมูลสำคัญในการ ทำนายข้อมูลที่เกี่ยวข้องกับคำพิพากษาในอนาคตได้ในการแยกคำให้การออกจากคำพิพากษา ทางคณะผู้วิจัยได้วิเคราะห์ข้อมูลและสังเกตเห็นลักษณะของส่วนของข้อมูลที่เป็นคำให้การว่ามีรูปแบบและตำแหน่งที่ชัดเจนตรงกันข้ามกับคำฟ้องที่มีรูปแบบ และการใช้คำที่หลากหลาย

จากนั้นคณะผู้วิจัยมีความเห็นว่าไม่จำเป็นต้องใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เพื่อตรวจหาคำให้การ แต่ได้ใช้นิพจน์ปรกติ (Regular Expression) ในการแยกรูปแบบส่วนเริ่มต้นของคำให้การออกจากคำให้การ Regular Expression เป็นเครื่องมือที่ใช้ในการสร้างแบบแผนของข้อความโดยใช้สัญลักษณ์พิเศษ เช่น Kleene star (*) ซึ่งหมายความว่ามีตัวอักษรนั้นอย่างน้อยหนึ่งตัว หรือ [a-z] ซึ่งหมายความว่าเป็นอักษรใดก็ได้ที่อยู่ในช่วง a-z Regular Expression เป็นเครื่องมือที่สำคัญในวิทยาการคอมพิวเตอร์และใช้ในการจัดการข้อมูลที่เป็นตัวอักษร สำหรับการตรวจหาคำให้การ โดยกำหนดให้ Regular Expression เป็นดังนี้ (ภาพ 16)

```
^(จำเลย)[ก-๙0-9. ]*(ให้การ)[ก-๙0-9. ]*ว่า
```

*รูปภาพ 16 แสดง Regular Expression ของการแยกคำให้การ*

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จาก Regular Expression ที่ถูกสร้างขึ้นรูปแบบของ paragraph ที่เป็นคำให้การจะต้องขึ้นต้นด้วยคำว่า “จำเลย” และ/หรือ ตามด้วยกลุ่มคำบางอย่างที่มีตัวอักษรประกอบด้วย “ก-๙0-9. “และจะตามด้วย คำว่า “ให้การ” หลังจากนั้นก็จะมีกลุ่มคำบางอย่างที่มีตัวอักษรประกอบด้วย “ก-๙0-9. “อีกครั้งและจะจบด้วยคำว่า “ว่า” ยกตัวอย่างเช่น (ภาพที่ 17)

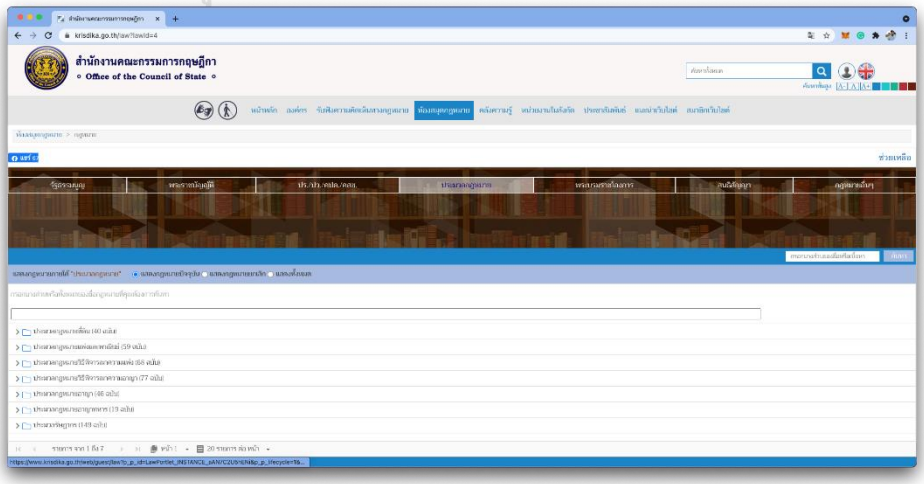
- จำเลยให้การว่า
- จำเลยทั้ง ๓ ให้การว่า
- จำเลยทั้ง 3 ให้การว่า
- จำเลยให้การยอมรับว่า

*รูปภาพ 17 แสดง Regular Expression ที่สามารถค้นหาและแยกข้อความได้*

ผู้วิจัยได้ตรวจสอบสุ่มผลการใช้ regular expression ดังกล่าว ผลปรากฏว่า จากการสุ่มตัวอย่างคำพิพากษาส่วนหนึ่ง (1000 คำพิพากษา)หากเป็นกรณีที่มีคำให้การอยู่ วิธีการในการแยกคำให้การดังกล่าวจะสามารถแยกแยะคำให้การได้อย่างสมบูรณ์แบบ จึงอาจสรุปได้ว่ารูปแบบของการบันทึกคำให้การมีลักษณะตายตัว กรณีดังกล่าวจึงไม่มีความจำเป็นต้องใช้การเรียนแบบมีผู้สอน (supervised learning) ในการตรวจแยกคำให้การออกจากเนื้อหาของคำพิพากษา

### 3.3.3 ข้อมูลนำเข้าในส่วนตัวบทบัญญัติของกฎหมาย

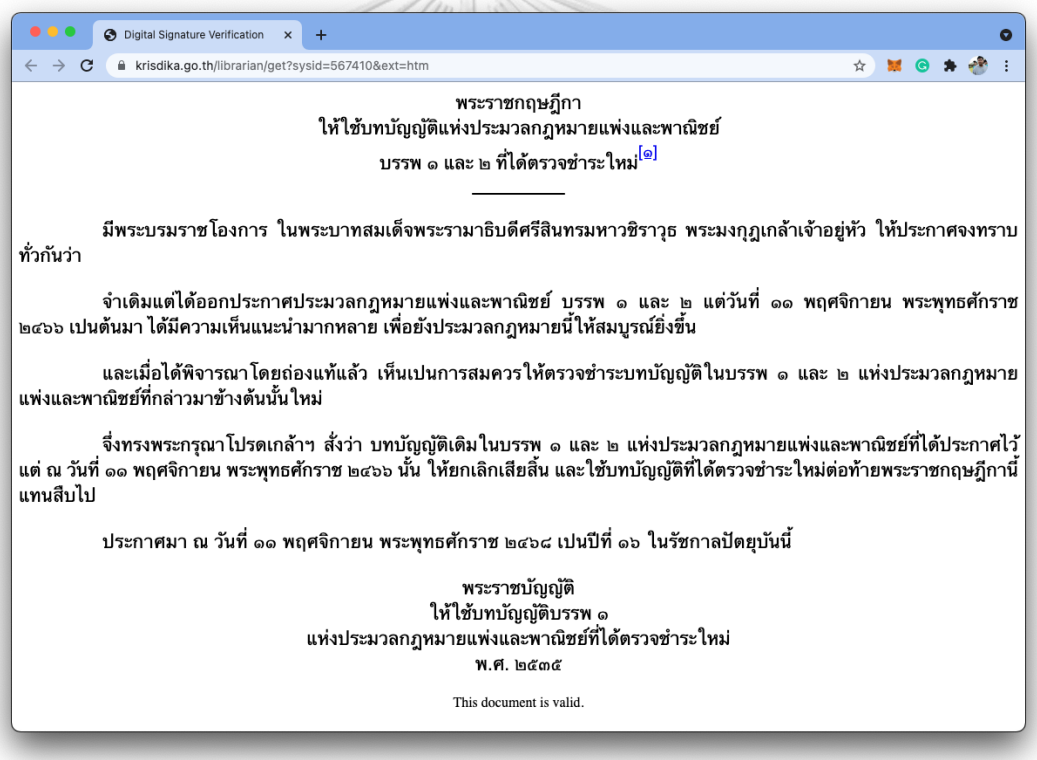
เพื่อรวบรวมข้อมูลประมวลกฎหมาย, คณะผู้วิจัยได้นำข้อมูลจากเว็บไซต์สำนักงานคณะกรรมการกฤษฎีกา (https://www.krisdika.go.th/) โดยใช้เทคนิค Web Scraping ซึ่งทำโดยการพัฒนาโปรแกรมด้วยภาษา Typescript และใช้ไลบรารี PhantomJS (https://phantomjs.org/) เพื่อจำลองการใช้งานเว็บไซต์ผ่านทางเบราว์เซอร์ Chromium โปรแกรมที่พัฒนาขึ้นรับข้อมูลต้นฉบับเป็นหน้าค้นหาข้อมูลกฎหมายของเว็บไซต์สำนักงานคณะกรรมการกฤษฎีกา จากนั้นระบบจะทำการจำลองการกดค้นหาข้อมูลกฎหมาย และเว็บไซต์จะแสดงผลลัพธ์ของการค้นหาดังรูปภาพที่ 18




*รูปภาพ 18 หน้าจอแสดงผลการค้นหาประมวลกฎหมายจากเว็บไซต์สำนักงานคณะกรรมการกฤษฎีกา*

หลังจากนั้นโปรแกรมจะทำการค้นหาลิงก์ของกฎหมายฉบับล่าสุด เช่น "ประมวลกฎหมายแพ่งและพาณิชย์ (ฉบับ Update ล่าสุด)" ซึ่งจะปรากฏในรูปภาพที่ 19 เมื่อพบลิงก์ดังกล่าว ระบบจะทำการบันทึกไฟล์เป็นรูปแบบ PDF เพื่อใช้ในการตรวจสอบความถูกต้องของข้อมูล จากนั้นโปรแกรมจะเข้าถึงหน้าหลักของ "ประมวลกฎหมายแพ่งและพาณิชย์ (ฉบับ Update ล่าสุด)" ดังแสดงในรูปภาพที่ 3.3 และทำการบันทึกไฟล์ HTML ของหน้านั้น และส่งข้อความให้ขั้นตอนถัดไปเพื่อสกัดข้อมูลออกมาในรูปแบบโครงสร้างตามที่แสดงในตารางที่ 3.1

*รูปภาพ 19 ลิงก์ของ ประมวลกฎหมายแพ่งและพาณิชย์ (ฉบับ Update ล่าสุด)*




*รูปภาพ 20 หน้าหลักของประมวลกฎหมายแพ่งและพาณิชย์ (ฉบับ Update ล่าสุด)*

*ตาราง 2 โครงสร้างการเก็บข้อมูลตัวบทบัญญัติของกฎหมาย*

| ชื่อ | คำอธิบาย |
|---|---|
| law_name | ชื่อของกฎหมาย เช่น ประมวณกฎหมายแพ่งและพาณิชย์ |
| section | เลขของมาตรา เช่น 420 |

| | |
|---|---|
| sub_section | เลขของมาตราย่อย เช่น มาตรา 193/1 sub_section ก็จะเป็น 1 |
| content | เนื้อหาของข้อกฎหมายนั้นๆ |

**3.3 รายละเอียดการทดลอง**

เมื่อฐานข้อมูลสำหรับการฝึกแบบจำลองถูกสร้างขึ้นแล้ว ขั้นตอนถัดไปเป็นการเตรียมข้อมูลเพื่อนำเข้าแบบจำลองและการออกแบบสถาปัตยกรรมแบบจำลองตามรูปแบบที่ใช้ในการพิสูจน์สมมติฐาน ดังนี้:

**3.3.1 การจัดเตรียมชุดข้อมูล**

หลังจากสร้างฐานข้อมูลแล้ว คณะผู้วิจัยจะแบ่งข้อมูลออกเป็นประโยคและตัดคำโดยใช้เครื่องมือ attacut ซึ่งเป็นเครื่องมือตัดคำอัตโนมัติในภาษาไพทอน ซึ่งความแม่นยำในการตัดคำของเครื่องมือชุดนี้มีค่ามากกว่า 90% (Chormai, Prasertsom et al. 2019) จากนั้นข้อความจะถูกแปลงเป็นค่าสัญญาณต่าง ๆ โดยการดำเนินการต่อกับข้อมูลคำฟ้อง คำให้การและคำอธิบายมาตราในฐานข้อมูล

ตัวอย่างข้อมูลคำฟ้อง

โจทก์ฟ้องว่าจำเลยที่ 1 ขณะปฏิบัติงานในทางการที่จ้างของจำเลยที่ 2 และที่ 3 ผู้เป็นนายจ้างได้ขับรถยนต์บรรทุกหกล้อหมายเลขทะเบียน 80-0283 สงขลาไปตามถนนสายเอเชียโดยความประมาทเลินเล่อปราศจากความระมัดระวังแซงรถยนต์บรรทุกสิบล้อซึ่งแล่นอยู่ข้างหน้าแล้วขับหลบไปทางด้านขวามือของถนนล้ำเข้าไปในช่องเดินรถของรถยนต์บรรทุกสิบล้อที่แล่นสวนทางมาเป็นเหตุให้รถยนต์บรรทุกของจำเลยที่ 1 พุ่งเข้าชนรถยนต์บรรทุกสิบล้อได้รับความเสียหายจากนั้นรถยนต์บรรทุกคันที่จำเลยที่1ขับเสียการทรงตัวส่ายไปมาและแล่นเข้าชนรถยนต์บรรทุกสี่ล้อหมายเลขทะเบียนน-9105 สงขลาของโจทก์ที่ 1 ซึ่งจอดอยู่ด้านซ้ายสุดของถนนฝั่งตรงกันข้ามได้รับความเสียหายพังยับเยินทั้งคันนอกจากนี้ยังทำให้สินค้าวัสดุก่อสร้างรวมหลายรายการของนายเนื่องจันทขวัญลูกค้าของโจทก์ที่ 2 ที่ได้สั่งซื้อสินค้าดังกล่าวไปจากโจทก์ที่ 2 ได้รับความเสียหายทั้งหมดซึ่งขณะเกิดเหตุอยู่ในระหว่างที่โจทก์ที่ 2 บรรทุกสินค้าวัสดุก่อสร้างดังกล่าวเพื่อนำไปส่งให้แก่นายเนื่องทำให้โจทก์ที่ 2 ต้องรับผิดชดใช้ค่าสินค้าให้แก่ทายาทของนายเนื่องตามสัญญาโจทก์ที่ 2 จึงเข้ารับช่วงสิทธิที่จะไล่เบี้ยเรียกร้องค่าสินค้าในส่วนที่โจทก์ที่ 2 ชำระไปแล้วการกระทำของจำเลยที่ 1 ทำให้โจทก์ทั้งสองเสียหายขอให้บังคับจำเลยทั้งสามร่วมกันชำระเงินแก่โจทก์ที่ 1 จำนวน 185,625 บาทพร้อมดอกเบี้ยในอัตราร้อยละ 7.5 ต่อปีจากต้นเงิน 180,000 บาทนับแต่วันฟ้องเป็นต้นไปจนกว่าจะชำระเสร็จและชำระเงินให้โจทก์ที่ 2 จำนวน 32,278 บาทพร้อมดอกเบี้ยในอัตราร้อยละ

ตัวอย่างข้อมูลคำให้การ

จำเลยที่ 1 ให้การว่า โจทก์ไม่เคยตกลงว่าจ้างจำเลยที่ 1 เป็นทนายความแก้ต่างให้ จำเลยที่ 1 เข้าเป็นทนายความในคดีแพ่งหมายเลขแดงที่ 1348/2531 ของศาลชั้นต้น ในฐานะลูกจ้างของบริษัทบัวหลวงประกันภัย จำกัด ซึ่งต้องกระทำตามสัญญาจ้างและตามคำสั่งของบริษัทบัวหลวงประกันภัย จำกัด และผู้บังคับบัญชาหาได้กระทำเพื่อสินจ้างหรือค่าจ้างตามที่โจทก์อ้างไม่ จำเลยที่ 1ซึ่งเป็นตัวแทนกระทำภายในขอบอำนาจของตัวการ โจทก์กับจำเลยที่ 1ไม่มีนิติสัมพันธ์ในลักษณะจ้างทำของ จำเลยที่ 1 ไม่ต้องรับผิดต่อโจทก์ ในเบื้องต้นจำเลยที่ 1 ได้เข้าเป็นทนายความให้โจทก์ตามหน้าที่ในฐานะลูกจ้างของบริษัทบัวหลวงประกันภัย จำกัด แต่ต่อมาจำเลยที่ 1 ได้ลาออกจากบริษัทบัวหลวงประกันภัย จำกัดแล้ว จำเลยที่ 1 จึงพ้นจากหน้าที่และความรับผิดชอบ จำเลยที่ 2 ในฐานะทนายความของบริษัทบัวหลวงประกันภัย จำกัด ได้เข้าเป็นทนายความรับผิดชอบแทนเหตุที่เกิดขึ้นจำเลยที่ 1 ไม่ทราบมาก่อนจำเลยที่ 1 ไม่ต้องรับผิดต่อโจทก์ จำเลยที่ 1 ไม่ได้ละทิ้งหน้าที่และความรับผิดชอบ มิได้กระทำการใด ๆ ด้วยความประมาทเลินเล่อในวันนัดพิจารณาและสืบพยานจำเลยที่ 2 ในคดีแพ่งหมายเลขแดงที่1348/2531 ของศาลชั้นต้น จำเลยที่ 2 ได้เข้าดำเนินกระบวนพิจารณาแล้ว ศาลชั้นต้นหาได้มีคำสั่งให้จำเลยที่ 2 ขาดนัดพิจารณาตามที่โจทก์อ้างไม่ จำเลยทั้งสองจึงไม่ต้องรับผิดต่อโจทก์ การที่จำเลยทั้งสองไม่เรียกบริษัทบัวหลวงประกันภัย จำกัด เข้าเป็นจำเลยร่วมหาใช่เพราะจำเลยทั้งสองละเลยต่อหน้าที่ไม่ แต่เนื่องจากโจทก์ไม่ใช่ผู้เอาประกันภัยตามสัญญาประกันภัย จึงไม่อาจเรียกบริษัทบัวหลวงประกันภัย จำกัด เข้าเป็นจำเลยร่วมได้ ที่โจทก์ต้องเสียหายต้องชดใช้ค่าเสียหายในคดีดังกล่าวเพราะลูกจ้างของโจทก์กระทำละเมิด และที่โจทก์อ้างว่าการกระทำของจำเลยที่ 1 ทำให้บริษัทบัวหลวงประกันภัย จำกัด หลุดพ้นความรับผิดตามสัญญาประกันภัย ทำให้โจทก์เสียหาย ก็ไม่ปรากฏว่าโจทก์ได้ฟ้องร้องเอาแก่บริษัทบัวหลวงประกันภัย จำกัด และไม่กระทำการใดเพื่อบรรเทาความเสียหาย ความเสียหายที่โจทก์ได้รับเกิดจากการกระทำ

ตัวอย่างข้อมูลคำอธิบายประมวลกฎหมายวิธีพิจารณาความแพ่ง มาตรา 55

เมื่อมีข้อโต้แย้งเกิดขึ้น เกี่ยวกับสิทธิหรือหน้าที่ของบุคคลใดตามกฎหมายแพ่ง หรือบุคคลใดจะต้องใช้สิทธิทางศาล บุคคลนั้นชอบที่จะเสนอคดีของตนต่อศาลส่วนแพ่งที่มีเขตอำนาจได้ ตามบทบัญญัติแห่งกฎหมายแพ่งและประมวลกฎหมายนี้

หลังจากตัดคำแล้วคณะผู้วิจัยได้ทำการวิเคราะห์ส่วนของคำพิพากษาที่สำคัญเพื่อนำไปใช้ในการปรับปรุงแบบจำลอง โดยการวิเคราะห์ได้แบ่งเป็น 2 ส่วนดังนี้:

**การวิเคราะห์ความยาวของข้อมูลคำพิพากษาและเอกสารทางกฎหมาย:** คณะผู้วิจัยทำการวิเคราะห์ความยาวของข้อมูลคำพิพากษาและเอกสารทางกฎหมายเพื่อตรวจสอบลักษณะและความหลากหลายของข้อมูลที่มีอยู่ในฐานข้อมูล โดยสำรวจจำนวนโทเคนในแต่ละเอกสาร เช่น จำนวนโทเคนที่มากที่สุด จำนวนโทเคนที่น้อยที่สุด และจำนวนโทเคนโดยเฉลี่ยของแต่ละเอกสาร ซึ่งจะแสดงในตารางที่ 3

**การตรวจสอบความผิดปกติของข้อมูล:** คณะผู้วิจัยนำโทเคนของคำพิพากษาและเอกสารทางกฎหมายมาวิเคราะห์เพื่อตรวจจับความผิดปกติของข้อมูล หากพบว่ามีความผิดปกติ เช่น โทเคนที่ขาดหายไปหรือเพิ่มเติมจากปกติ คณะผู้วิจัยจะทำการปรับปรุงและนำข้อมูลเหล่านี้ไปใช้ในการพัฒนาแบบจำลองให้มีประสิทธิภาพมากยิ่งขึ้น

ขั้นตอนเหล่านี้จะช่วยให้คณะผู้วิจัยเข้าใจลักษณะของข้อมูลและความหลากหลายในฐานข้อมูล เพื่อนำข้อมูลเหล่านี้ไปใช้ในการพัฒนาและปรับปรุงแบบจำลองให้มีประสิทธิภาพในการทำนายคำพิพากษามากยิ่งขึ้น

*ตาราง 3 จำนวนโทเคนในข้อมูลรับเข้าแต่ละประเภท*

| ประเภทข้อมูล | จำนวนโทเคนที่มากที่สุด | จำนวนโทเคนที่น้อยที่สุด | จำนวนโทเคนโดยเฉลี่ย |
|---|---|---|---|
| คำฟ้อง | 2,827 | 3 | 180.62 |
| คำให้การ | 1,123 | 1 | 72.67 |
| คำอธิบายมาตรา | 389 | 21 | 105.09 |

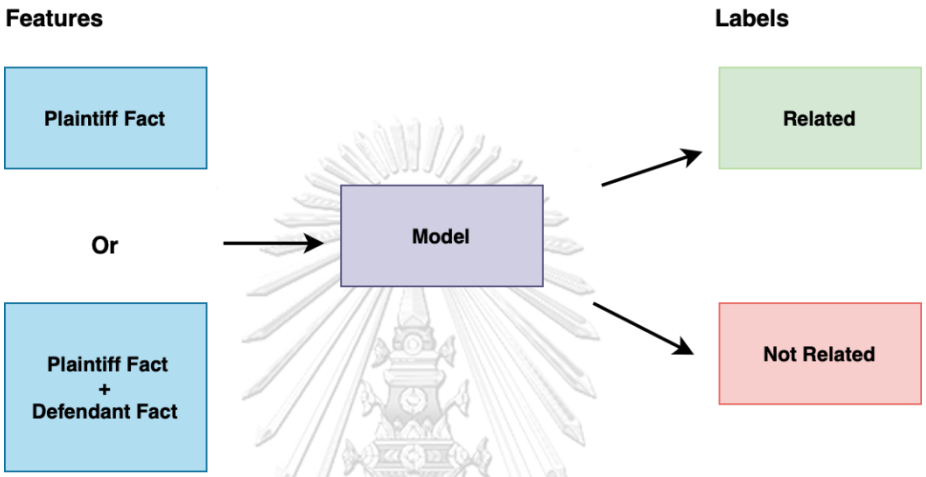
CHULALONGKORN UNIVERSITY

จากการวิเคราะห์ตารางที่ 3 พบว่าคำฟ้องมีจำนวนโทเคนมากที่สุดเมื่อเปรียบเทียบกับคำให้การและคำอธิบายมาตรา ซึ่งนักวิจัยพบว่าคำฟ้องมักจะกล่าวถึงรูปการณ์ของจำเลยมากกว่า 1 คน ในขณะที่คำให้การมักจะกล่าวถึงกฎหมายจากหน่วยงานอื่น ๆ เช่น กฎหมายกระทรวง และมีการกล่าวถึงพยานและหลักฐานที่ยืนยันว่าจำเลยไม่ผิด เชื่อมั่นได้ว่าการวิเคราะห์ข้อมูลเหล่านี้จะช่วยให้แบบจำลองมีความเข้าใจและความสามารถในการทำนายคำพิพากษาได้อย่างแม่นยำและมีประสิทธิภาพมากยิ่งขึ้น

### 3.3.2 รูปแบบสถาปัตยกรรมในการทดลอง

แบบจำลองแต่ละแบบที่สร้างจะถูกนำมาทดลองในสถาปัตยกรรม 2 รูปแบบคือ รูปแบบการจำแนกข้อมูลสองประเภท (Binary Classification) และรูปแบบสถาปัตยกรรมโครงข่ายประสาทเทียมสยาม (Siam Neural Network) เพื่อทดสอบสมมติฐานที่ต่างกัน โดยมีรายละเอียด ดังนี้

การจำแนกข้อมูลสองประเภท (Binary Classification)




*รูปภาพ 21 แผนภาพแบบจำลองแบบจำแนกข้อมูลสองประเภท*

ในการจำแนกข้อมูลสองประเภทดังกล่าว แบบจำลองที่ได้มีบทบาทเป็นตัวแปลง (Encoder) ที่ใช้ในการแปลงข้อมูลเข้าเป็นเวกเตอร์ค่าสัญญาณที่สามารถนำไปใช้ในการทำนายต่อได้ โดยผลลัพธ์ที่ต้องการคือป้ายกำกับ (Label) ที่ระบุว่าส่วนของข้อเท็จจริงในคำพิพากษานั้น "เกี่ยวข้อง" หรือ "ไม่เกี่ยวข้อง" กับมาตราต่าง ๆ นอกเหนือจากนี้ คณะผู้วิจัยต้องการทดสอบสมมติฐานที่ว่าการเพิ่มข้อมูลคำให้การเข้าร่วมกับคำฟ้องจะช่วยให้แบบจำลองสามารถทำนายมาตราที่ถูกต้องได้ดีขึ้น ดังนั้น ข้อมูลเข้าในสถาปัตยกรรมนี้จะแบ่งออกเป็น 2 ลักษณะคือ

**การฝึกแบบจำลองโดยใช้เฉพาะข้อเท็จจริงส่วนคำฟ้อง (plaint)** เป็นข้อมูลเข้าในกระบวนการฝึกแบบจำลอง โดยใช้คำฟ้องเพียงอย่างเดียวเป็นข้อมูลที่นำเข้า

**การฝึกแบบจำลองที่ใช้ข้อเท็จจริงทั้งส่วนคำฟ้อง (plaint) และคำให้การ (defense)** เป็นข้อมูลเข้าในกระบวนการฝึกแบบจำลอง ซึ่งประกอบไปด้วยทั้งคำฟ้องและคำให้การ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

โดยการแบ่งข้อมูลเข้าเป็นสองลักษณะเพื่อเป็นส่วนหนึ่งของการออกแบบสถาปัตยกรรมแบบจำลอง เพื่อให้แบบจำลองมีความสามารถในการทำนายมาตราที่ถูกต้องและปรับปรุงแบบจำลองได้ให้มีประสิทธิภาพมากยิ่งขึ้น

***3.3.2.1 ใช้คำฟ้องเท่านั้นเป็นค่าสัญญาณ***

ตัวแปลง(encoder) ที่ใช้ ได้แก่ แบบจำลอง Bi-LSTM, CNN และ BERT และใช้ข้อมูลรับเข้าเป็นคำฟ้องเพียงอย่างเดียว ส่วนผลการทำนายของแบบจำลองจะเป็นการระบุว่าคำฟ้องนั้น “เกี่ยวข้อง” หรือ “ไม่เกี่ยวข้อง” กับมาตรานั้น ๆ

*ตาราง 4 ตารางแสดงตัวอย่างรูปแบบข้อมูลของแบบจำลองที่ใช้เฉพาะคำฟ้องเป็นค่าสัญญาณ*

| ID ของคำพิพากษา | คำฟ้อง | รหัสแทนมาตราที่เกี่ยวข้อง | ป้ายกำกับ |
|---|---|---|---|
| 1 | …. | 251 | 1 |
| 2 | …. | 629 | 0 |
| 3 | …. | 251 | 1 |

จากตารางที่ 4 ข้อมูลเริ่มต้นที่ใช้ในการฝึกแบบจำลองประกอบด้วยเลขของคำพิพากษา (ID), คำฟ้อง (Plaint), และรหัสแทนมาตราที่เกี่ยวข้อง เมื่อนำมาฝึกแบบจำลอง จะเพิ่มส่วนป้ายกำกับเข้าไปเป็นผลลัพธ์ในกระบวนการฝึก โดยใช้ค่า 0 และ 1 ในการแทนผลลัพธ์ โดยที่ในที่นี้ได้ทำการสมมติว่าชุดข้อมูลประกอบด้วย 3 คดีความเป็นตัวอย่างเฉพาะสำหรับมาตรารหัส 251 โดยถ้าคำฟ้องใดเกี่ยวข้องกับมาตรารหัส 251 จะมีป้ายกำกับเป็น 1 และหากไม่เกี่ยวข้องจะมีป้ายกำกับเป็น 0 แบบจำลองที่เกี่ยวข้องกับมาตรารหัสอื่น ๆ ก็จะเป็นไปตามลักษณะเดียวกัน

***3.3.2.2 ใช้คำฟ้องและคำให้การเป็นค่าสัญญาณ***

ตัวแปลงที่ใช้ ได้แก่ แบบจำลอง Bi-LSTM, CNN, BERT และใช้ข้อมูลรับเข้าเป็นทั้งคำฟ้องและคำให้การ ส่วนผลการทำนายของแบบจำลองจะเป็นการระบุว่าคำฟ้องนั้น “เกี่ยวข้อง” หรือ “ไม่เกี่ยวข้อง” กับมาตรานั้น ๆ

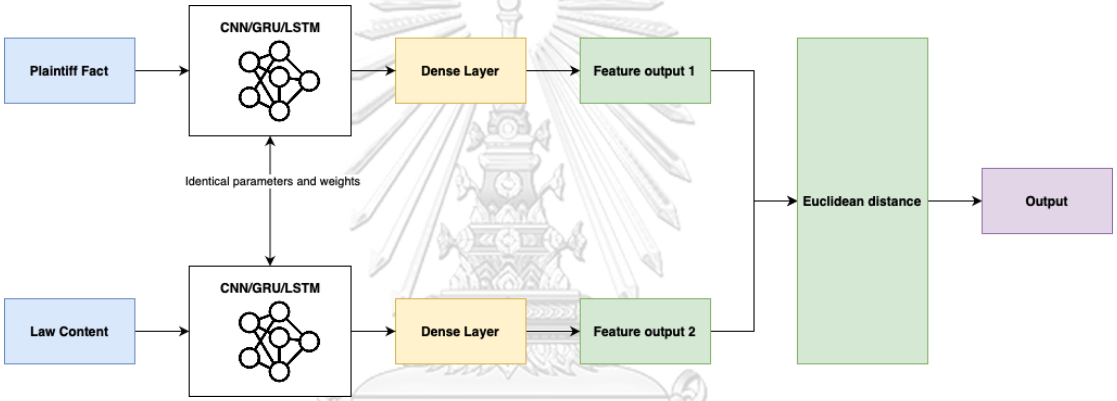
*ตาราง 5 ตารางแสดงตัวอย่างรูปแบบข้อมูลของแบบจำลองที่ใช้ทั้งคำฟ้องและคำให้การเป็นค่าสัญญาณ*

| ID ของคำพิพากษา | คำฟ้อง | คำให้การ | รหัสแทนมาตราที่เกี่ยวข้อง | ป้ายกำกับ |
|---|---|---|---|---|
| 1 | …. | …. | 251 | 1 |
| 2 | …. | …. | 629 | 0 |

| | | | | |
|---|---|---|---|---|
| 3 | …. | …. | 251 | 1 |

จากตารางที่ 5 ข้อมูลเริ่มต้นที่ใช้ในการฝึกแบบจำลองประกอบด้วยเลขของคำพิพากษา (ID), คำฟ้อง (Plaint), คำให้การ (defense), และรหัสแทนมาตราที่เกี่ยวข้อง เมื่อนำมาฝึกแบบจำลอง จะเพิ่มส่วนป้ายกำกับเข้าไปเป็นผลลัพธ์ในกระบวนการฝึก โดยใช้ค่า 0 และ 1 ในการแทนผลลัพธ์ โดยที่ในงานวิจัยนี้ได้สมมติว่าชุดข้อมูลประกอบด้วย 3 คดีความเป็นตัวอย่างเฉพาะสำหรับมาตรารหัส 251 โดยถ้าคำฟ้องใดเกี่ยวข้องกับมาตรารหัส 251 จะมีป้ายกำกับเป็น 1 และหากไม่เกี่ยวข้องจะมีป้ายกำกับเป็น 0 แบบจำลองที่เกี่ยวข้องกับมาตรารหัสอื่น ๆ ก็จะเป็นไปตามลักษณะเดียวกัน

***3.3.2.3 สถาปัตยกรรมโครงข่ายประสาทเทียมสยาม (Siam Neural Network)***




*รูปภาพ 22 แผนภาพแบบจำลองแบบโครงข่ายประสาทเทียมสยาม*

เนื่องจากมีคำพิพากษาที่รวบรวมมามากมาย แต่มีการกระจายตัวของแต่ละมาตราที่ไม่สมดุลกัน บางมาตรามีจำนวนมากกว่ามาตราอื่น ๆ และแบบจำลองรูปแบบการจำแนกข้อมูลสองประเภทมีความสามารถจำกัดที่ทำนายได้เฉพาะมาตราที่ผ่านการฝึกฝนเท่านั้น ถ้ามีข้อมูลมาตราน้อย แบบจำลองจะมีประสิทธิภาพลดลงไป ดังนั้น คณะผู้วิจัยจึงรวบรวมคำอธิบายเพิ่มเติมของมาตราเพื่อฝึกแบบจำลองรูปแบบโครงข่ายประสาทเทียมสยาม แบบการเรียนรู้ด้วยข้อมูลจำนวนน้อย (Few-shot Learning) และแบบการเรียนรู้โดยไม่มีข้อมูลปฐมภูมิ (Zero-shot Learning) โดยแบบจำลองรูปแบบโครงข่ายประสาทเทียมสยามเหล่านี้เหมาะกับการทำนายข้อมูลที่มีข้อมูลชุดฝึกจำนวนน้อยหรือไม่มีข้อมูลชุดฝึกเลยมากกว่าแบบจำลองรูปแบบการจำแนกข้อมูลสองประเภท

สถาปัตยกรรมโครงข่ายประสาทเทียมสยาม (Siamese Neural Network) เป็นลักษณะแบบจำลองที่ใช้ค่าน้ำหนักร่วมกันในการประมวลผล โดยสามารถป้อนข้อมูล

ลงไปมากกว่า 1 อย่างได้ ถ้าข้อมูลเหล่านั้นมีลักษณะคล้ายกันมาก ค่าสัญญาณที่ได้จากแบบจำลองก็จะมีความคล้ายเช่นกัน คณะผู้วิจัยใช้แนวคิดของแบบจำลองนี้เพื่อตัดสินว่าข้อเท็จจริงเกี่ยวข้องกับมาตราที่ให้หรือไม่ โดยให้แบบจำลองเปรียบเทียบความคล้ายระหว่างข้อความในข้อเท็จจริงกับคำอธิบายของมาตรา ด้วยการอ้างอิงสมมติฐานว่าหากข้อเท็จจริงเกี่ยวกับมาตราดังกล่าว ความหมายและลักษณะคำควรจะมีความคล้ายคลึงกัน ผลลัพธ์ที่ได้จากแบบจำลองนี้คือป้ายกำกับที่ระบุว่าส่วนของข้อเท็จจริงในคำพิพากษา "เกี่ยวข้อง" หรือ "ไม่เกี่ยวข้อง" กับมาตราต่าง ๆ เช่นเดียวกัน ดังแสดงในรูปที่ 22 ในการทดลอง ใช้แบบจำลองโครงข่ายย่อย (Sub-network) 2 รูปแบบที่แตกต่างกัน ได้แก่ โครงข่ายที่ใช้ข้อมูลจำนวนน้อยในการเรียนรู้ (Few-shot Network) และโครงข่ายที่ไม่มีข้อมูลปฐมภูมิในการเรียนรู้ (Zero-shot Network) ซึ่งมีรายละเอียดต่างกัน

**โครงข่ายที่ใช้ข้อมูลจำนวนน้อยในการเรียนรู้ (Few-shot Network)** จะมีการใช้แบบจำลองต่างๆ ที่มาเป็นส่วนประกอบของสถาปัตยกรรมโครงข่ายประสาทเทียมสยาม คณะผู้วิจัยจะฝึกแบบจำลองหนึ่งเพื่อสังเกตลักษณะค่าสัญญาณของคำฟ้อง และฝึกแบบจำลองอีกตัวหนึ่งเพื่อสังเกตลักษณะค่าสัญญาณของคำอธิบายมาตรา จากนั้นค่าสัญญาณที่ได้จากแบบจำลองทั้งสองจะถูกนำมาเทียบค่าสูญเสียเปรียบต่าง (Contrastive Loss) เพื่อทำนายว่าคำฟ้องนั้นเกี่ยวข้องกับคำอธิบายมาตราที่ให้หรือไม่ อย่างไรก็ตามในกรณีของโครงข่ายย่อยแบบ Few-shot ข้อมูลที่มีป้ายกำกับตามที่กำหนดในช่วงการเรียนรู้ของแบบจำลองจะปรากฏในข้อมูลชุดทดสอบด้วย

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

*ตาราง 6 ตารางแสดงตัวอย่างรูปแบบข้อมูลของแบบจำลองในโครงข่ายย่อยแบบ Few-shot*

| ID ของคำพิพากษา | | คำฟ้อง | คำอธิบายมาตรา | รหัสแทนมาตราที่เกี่ยวข้อง | ป้ายกำกับ |
|---|---|---|---|---|---|
| 1 | | …. | A | 251 | 1 |
| 2 | | …. | B | 629 | 0 |
| 3 | | …. | A | 251 | 1 |

จากตารางที่ 6 ในข้อมูลแรกเริ่มประกอบด้วยเลข ID ของคำพิพากษา คำฟ้อง คำอธิบายมาตรา และรหัสแทนมาตราที่เกี่ยวข้อง เมื่อนำมาฝึกแบบจำลอง Few-shot Network จะเพิ่มส่วนป้ายกำกับเข้าไปเป็นผลลัพธ์ในการฝึก โดยแทนผลลัพธ์ด้วย 0 และ 1 ในที่นี้ ทางคณะผู้วิจัยได้ทำการสมมติว่าทั้งชุดข้อมูลประกอบด้วย 3 คดีความ และเป็นตัวอย่างเฉพาะแบบจำลองของมาตรารหัส 251 หากคู่คำฟ้องและคำอธิบายมาตราใดสอดคล้องกัน จะมีป้ายกำกับเป็น 1 หากไม่สอดคล้องกัน จะมีป้ายกำกับ

เป็น 0 แบบจำลองในมาตรารหัสอื่น ๆ ก็จะมีลักษณะเช่นนี้เช่นกัน ซึ่งทำให้ Few-shot Network มีแบบจำลองแยกกันไปเป็นแบบจำลองเฉพาะของแต่ละมาตรา รวมทั้งหมด 4 มาตรา

**โครงข่ายที่ปราศจากข้อมูลปฐมภูมิในการเรียนรู้ (Zero-shot Network)** จะใช้แบบจำลองต่าง ๆ เป็นตัวแปลงบนสถาปัตยกรรมโครงข่ายประสาทเทียมสยาม โดยคณะผู้วิจัยจะฝึกแบบจำลองหนึ่งให้สังเกตลักษณะค่าสัญญาณของคำฟ้อง และอีกแบบจำลองหนึ่งสังเกตลักษณะค่าสัญญาณของคำอธิบายมาตรา แล้วนำค่าสัญญาณจากแบบจำลองทั้งสองมาเทียบค่าสูญเสียเปรียบต่าง เพื่อทำนายว่าคำฟ้องนั้นมีความเกี่ยวข้องกับคำอธิบายมาตราที่ให้หรือไม่ แต่โครงข่ายย่อยแบบ Zero-shot แตกต่างจากโครงข่ายลักษณะแรกคือ แบบจำลองในโครงข่ายลักษณะนี้จะไม่มีโอกาสได้เห็นลักษณะข้อมูลที่มีป้ายกำกับตามที่กำหนดในช่วงการเรียนรู้ของแบบจำลองในข้อมูลชุดทดสอบเลย และต้องทำนายข้อมูลป้ายกำกับใหม่ที่เป็นมาตราที่ไม่เกี่ยวข้องกับมาตรา 420 ทำให้แบบจำลองไม่มีองค์ความรู้ของลักษณะข้อมูลดังกล่าวมาก่อน

*ตาราง 7 ตารางแสดงตัวอย่างรูปแบบข้อมูลของแบบจำลองในโครงข่ายย่อยแบบ Zero-shot*

| ID ของคำพิพากษา | คำฟ้อง | คำอธิบายมาตรา | รหัสแทนมาตราที่เกี่ยวข้อง | ป้ายกำกับ |
|---|---|---|---|---|
| 1 | …. | A | 251 | 1 |
| 1 | …. | B | 629 | 0 |
| 1 | …. | C | 135 | 0 |
| 2 | …. | A | 251 | 0 |
| 2 | …. | B | 629 | 1 |
| 2 | …. | C | 135 | 0 |

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 7 ข้อมูลแรกประกอบด้วยเลข ID ของคำพิพากษา คำฟ้อง คำอธิบายมาตรา และรหัสแทนมาตราที่เกี่ยวข้อง เมื่อนำมาฝึกแบบจำลอง จะเพิ่มส่วนป้ายกำกับเข้าไปเป็นผลลัพธ์ในการฝึก โดยแทนผลลัพธ์ด้วย 0 และ 1 ในงานวิจัยนี้ได้ทำการสมมติว่าชุดข้อมูลประกอบด้วย 2 คดีความ และ 3 มาตราเท่านั้น แบบจำลองจะจับคู่คำฟ้องและคำอธิบายมาตราในทุกรูปแบบที่เป็นไปได้ ทำให้เกิดป้ายกำกับเป็น 0 มากขึ้น เปรียบเสมือนแบบจำลองนี้ต้องฝึกแยกแต่ละมาตรา เนื่องจากสามารถเรียนรู้ลักษณะความสอดคล้องหรือไม่สอดคล้องในทุกรูปแบบของข้อมูลที่มีอยู่ได้ ชุดข้อมูลลักษณะนี้ทำให้แบบจำลองสามารถเรียนรู้ลักษณะทุกมาตราในหนึ่งแบบจำลองโดยไม่

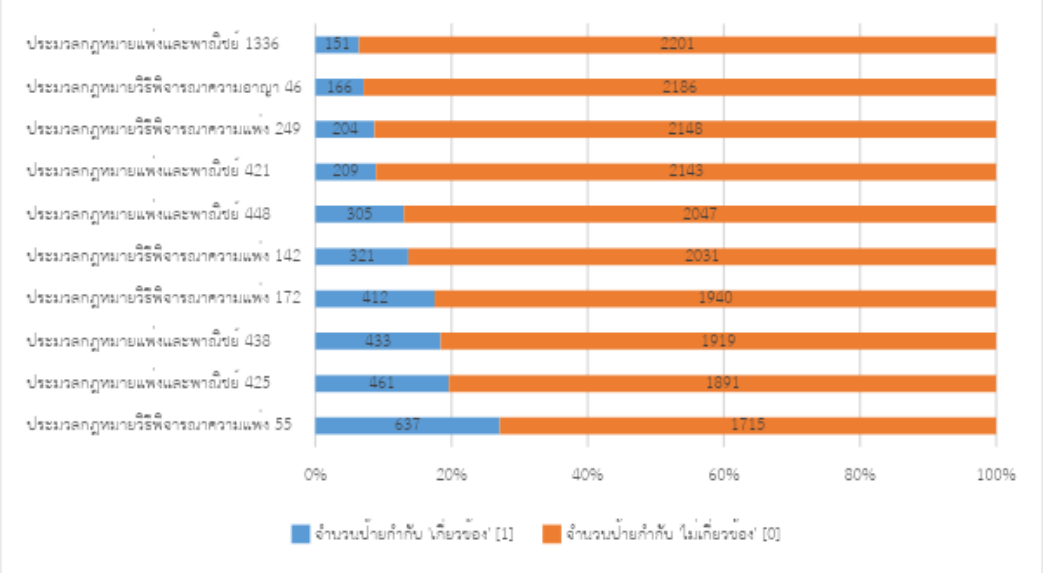
จำเป็นต้องฝึกแบบจำลองแยกแต่ละมาตรา ข้อมูลเพิ่มเติมเกี่ยวกับการจัดเตรียมข้อมูลและรายละเอียดอื่นๆ สามารถเรียนรู้ได้จากข้อมูลที่ผู้วิจัยได้แจกแจงไว้

### 3.3.3 การออกแบบการทดลอง

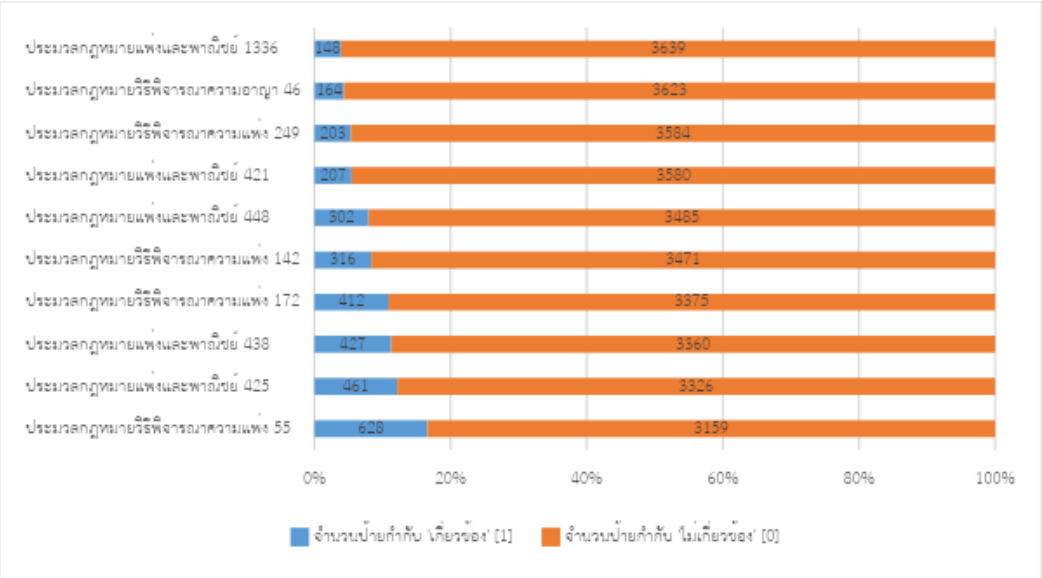
#### *3.3.3.1 การแบ่งข้อมูล*

ในส่วนของการฝึกแบบจำลอง คณะผู้วิจัยใช้ลักษณะแบบจำลองตามที่ได้กล่าวไปข้างต้น โดยแบ่งชุดข้อมูลออกเป็น 3 ส่วน ได้แก่ ข้อมูลชุดฝึก 70% ข้อมูลวัดเพื่อปรับแบบจำลองระหว่างฝึก 10%และข้อมูลชุดทดสอบประสิทธิภาพของแบบจำลอง 20%

จำนวนชุดข้อมูลที่ใช้ในในการฝึกฝนแบบจำลองการจำแนกข้อมูลสองประเภท เท่ากับ 2,352 คู่ชุดข้อมูลต่อหนึ่งมาตราที่เกี่ยวข้อง ตามรูปที่ *23* ในขณะที่แบบจำลองแบบโครงข่ายประสาทเทียมสยามโครงข่ายย่อยแบบการเรียนรู้ด้วยข้อมูลจำนวนน้อย (Few-shot Sub-network) และโครงข่ายย่อยแบบการเรียนรู้โดยปราศจากข้อมูลปฐมภูมิ (Zero-shot Sub-network) ล้วนใช้ข้อมูลตั้งต้นชุดเดียวกันจำนวน 15,055 ข้อมูล หลังจากปรับข้อมูลมีจำนวนชุดข้อมูลที่ใช้ในการฝึกฝนทั้งหมดเท่ากับ 37,870 คู่ชุดข้อมูล ซึ่งประกอบด้วยสัดส่วนของป้ายกำกับของแต่ละมาตราตามรูปที่ *2* *4*




*รูปภาพ 23 สัดส่วนป้ายกำกับในแบบจำลองการจำแนกแบบสองประเภทของแต่ละมาตรา*




*รูปภาพ 24 สัดส่วนป้ายกำกับในแบบจำลองแบบโครงข่ายประสาทเทียมสยามของแต่ละมาตรา*

### *3.3.3.2 การปรับจูนไฮเปอร์พารามิเตอร์*

ในการฝึกแบบจำลองโครงข่ายประสาทเทียม คณะผู้วิจัยใช้ Word Embedding จำนวน 300 มิติที่สร้างจากไลบรารี fastText ซึ่งได้เรียนรู้ข้อมูลคำจากคลังข้อมูล Common Crawl และคลังข้อมูล Wikipedia โดยมีขั้นตอนดังนี้:

1. ปรับจำนวนรอบของการฝึกแบบจำลอง (Epoch) ตามความเหมาะสมของแบบจำลอง
2. ควบคุมขนาดการแบ่งชุดข้อมูลเพื่อป้อนเข้าแบบจำลองระหว่างฝึก (Batch Size) ให้อยู่ระหว่าง 24-60
3. ใช้อัลกอริทึมอดัม (Adam) เป็น Optimizer เพื่อหาค่าน้ำหนักที่เหมาะสมที่สุดที่ทำให้ค่าสูญเสียน้อยที่สุด

สำหรับแบบจำลอง CNN คณะผู้วิจัยได้กำหนดไฮเปอร์พารามิเตอร์เพิ่มเติมดังนี้:

1. ในแบบจำลองการจำแนกข้อมูลสองประเภท กำหนดจำนวนฟิลเตอร์ (Filter) เท่ากับ 128 และขนาดฟิลเตอร์ (Kernel) เท่ากับ 28
2. ส่วนแบบจำลองแบบโครงข่ายประสาทเทียมสยามกำหนดจำนวนฟิลเตอร์เท่ากับ 64 และ 80 และขนาดฟิลเตอร์เท่ากับ 5 และ 15
3. กำหนดขนาดเวกเตอร์สำหรับการเลือกค่าสูงสุดของเวกเตอร์ (pool size) เท่ากับ 2

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4. ใช้ฟังก์ชันกระตุ้น ReLU เป็นฟังก์ชันกระตุ้นในชั้นประมวลผล
5. ในชั้นประมวลผล มีการสุ่มปิดหน่วยประมวลผล (Dropout) ร้อยละ 50 เพื่อลดการเกิดปัญหา Overfitting ที่เกิดขึ้นในข้อมูลชุดฝึก

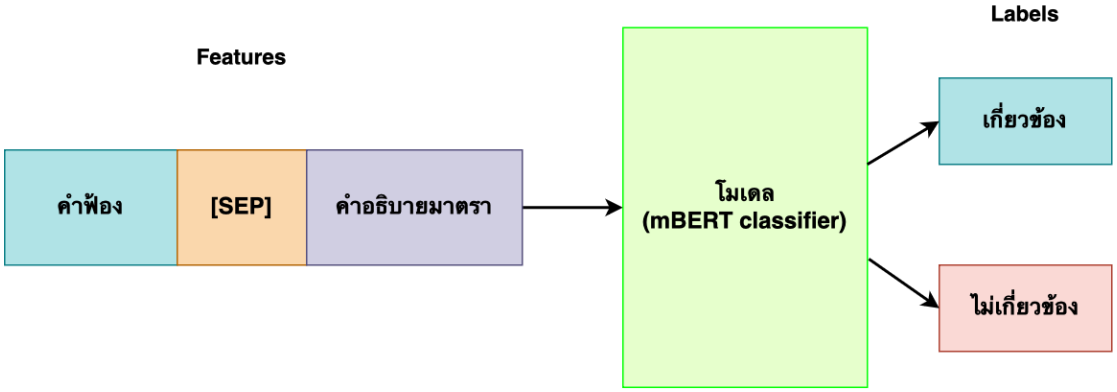
คณะผู้วิจัยยังใช้ Word Embedding จากแบบจำลอง BERT ชนิด Multilingual base model (หรือ mBERT) เพื่อรองรับภาษาต่างๆ รวมถึงภาษาไทยด้วย แบบจำลอง mBERT สามารถรองรับได้กว่า 104 ภาษา โดยมี Word Embedding ที่มีขนาด 768 มิติ และได้รับการเรียนรู้คำมาก่อนจากข้อมูลทั้งหมดในคลังข้อมูล Wikipedia ของทุกภาษา ทำให้ Word Embedding ของ mBERT มีข้อมูลคำที่มากกว่า Word Embedding อื่นๆ ที่เรียนรู้ข้อมูลคำมาก่อนแล้ว และมีแนวโน้มที่จะมีประสิทธิภาพมากกว่า

คณะผู้วิจัยทดลองใช้แบบจำลอง mBERT ในรูปแบบการจำแนกข้อมูลสองประเภท (Binary Classification) โดยใช้ mBERT เป็นตัวแปลง (Encoder) ข้อมูลเพื่อเปลี่ยนข้อมูลรับเข้าให้กลายเป็นค่าสัญญาณที่แบบจำลองสามารถนำไปทำนายต่อไปได้ แต่ข้อมูลรับเข้าของ BERT และ mBERT จะมีลักษณะต่างออกไป โดยข้อมูลรับเข้าจะมีเพียงอันเดียว แต่สามารถใส่ข้อมูลมากกว่า 2 อย่างได้ด้วยการใช้อักขระพิเศษ [SEP] เพื่อคั่นและเชื่อมเป็นข้อมูลรับเข้าในหน่วยเดียวกัน

ส่วนอีกรูปแบบหนึ่งคือโครงข่ายประสาทเทียมสยาม (Siamese Neural Network) ซึ่งในที่นี้จะใช้ Word Embedding ของ mBERT มาหาเทียบหาความคล้ายระหว่างข้อความในข้อเท็จจริงกับคำอธิบายมาตราแทน และใช้แบบจำลองอื่นมากำหนดเกณฑ์เพื่อตัดสินว่าข้อเท็จจริงนั้นเกี่ยวข้องกับคำอธิบายมาตราที่ให้หรือไม่

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1. mBERT แบบ Next sentence prediction (เทียบเคียงการจำแนกข้อมูลสองประเภท)

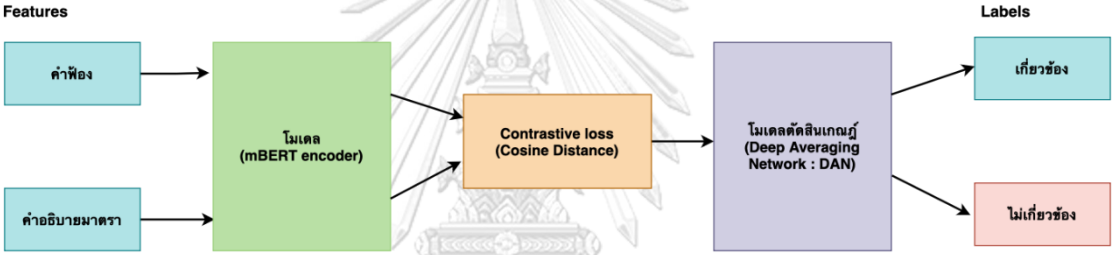



*รูปภาพ 25 แผนภาพแบบจำลอง mBERT แบบจำแนกข้อมูลสองประเภท*

แบบจำลองนี้ใช้ mBERT เป็นเครื่องจำแนกประเภทโดยตรง โดยรวมคำฟ้องและคำอธิบายมาตราให้เป็นหนึ่งข้อมูลป้อนเข้าด้วยเครื่องหมายคั่นพิเศษ [SEP] โดยแบ่งความยาวให้เป็นข้างละ 255 โทเคน ความยาวสูงสุดของข้อมูลป้อนเข้าจึงเป็น 510 โทเคน และเมื่อใส่ข้อมูลป้อนเข้าแล้วจะให้แบบจำลองจำแนกออกมาว่าคำฟ้องและคำอธิบายมาตรานี้มีความ ‘เกี่ยวข้อง’ หรือ ‘ไม่เกี่ยวข้อง’

การกำหนดไฮเปอร์พารามิเตอร์ ได้แก่ การควบคุมขนาดการแบ่งชุดข้อมูลเพื่อป้อนเข้าแบบจำลองระหว่างฝึก (Batch Size) ให้เป็น 10 และใช้อัลกอริทึมอดัม (Adam) สำหรับหาค่าน้ำหนักที่เหมาะสมที่สุดที่ทำให้ค่าสูญเสียน้อยที่สุด (Optimizer) โดยที่มีอัตราค่าการเรียนรู้ (Learning rate) อยู่ที่ 5e-6

2. mBERT แบบใช้ค่าระยะห่างโคไซน์ (Cosine distance) จาก Word Embeddings (เทียบเคียงโครงข่ายประสาทเทียมแบบสยาม)




*รูปภาพ 26 แผนภาพแบบจำลอง mBERT แบบโครงข่ายประสาทเทียมสยาม*

ในแบบจำลองนี้คณะผู้วิจัยได้นำคำฟ้องและคำอธิบายมาตรามาแปลงเป็น Word Embedding ด้วยตัวแปลง mBERT เพื่อนำค่าสัญญาณมาคำนวณหาค่าความต่างด้วยค่าระยะห่าง Euclidian Distance ระหว่างค่าสัญญาณคำฟ้องและค่าสัญญาณคำอธิบายมาตรา เมื่อได้ค่าความต่างระหว่างคำฟ้องและคำอธิบายมาตราแล้ว เท่ากับว่าได้มา 1 ค่าสัญญาณ ในขั้นตอนถัดไปคือการนำค่าสัญญาณ 1 มิตินี้เข้าแบบจำลองอื่น เพื่อให้แบบจำลองอื่นช่วยตัดสินใจว่า หากมีค่าความต่างเท่านี้ ควรให้ป้ายกำกับว่าคู่ชุดข้อมูล คำฟ้องและคำอธิบายมาตรา เกี่ยวข้องกันหรือไม่ โดยใช้โครงข่ายประสาทเทียมเฉลี่ยเชิงลึก (Deep Averaging Neural Network: DAN) เป็นเครื่องกำหนดเกณฑ์ดังกล่าว

สำหรับแบบจำลอง DAN คณะผู้วิจัยควบคุมขนาดการแบ่งชุดข้อมูลเพื่อป้อนเข้าแบบจำลองระหว่างฝึก (Batch Size) ให้อยู่ระหว่าง 24-64 และใช้อัลกอริทึมอดัม (Adam) สำหรับหาค่าน้ำหนักที่เหมาะสมที่สุดที่ทำให้ค่าสูญเสียน้อยที่สุด (Optimizer)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

# บทที่ 4
# ผลการวิจัย

ในบทนี้จะอธิบายผลลัพธ์ที่ได้จากการทดลองเพื่อประเมินประสิทธิภาพของแบบจำลองในการทำนายมาตราที่เกี่ยวข้อง โดยในการทดลองนั้นจะเน้นการทำนายความเกี่ยวข้องระหว่างคำฟ้องและมาตรา ผลการทดลองจะแสดงให้เห็นถึงประสิทธิภาพของแบบจำลองแบบการจำแนกข้อมูลสองประเภทที่ใช้ข้อมูลจำนวนน้อยในการฝึกฝน รวมถึงผลลัพธ์ที่ได้จากแบบจำลองแบบโครงข่ายประสาทเทียมสยามที่ทดสอบความสามารถในการทำนายมาตราที่เกี่ยวข้องโดยใช้ชุดข้อมูลฝึกที่จำกัด นอกจากนี้ยังจะมีการนำแบบจำลองโครงข่ายประสาทเทียมสยามมาประยุกต์ใช้ในระบบค้นคืนมาตราที่เกี่ยวข้อง ผลการทดลองจะแสดงประสิทธิภาพของการใช้แบบจำลองโครงข่ายประสาทเทียมสยามในการค้นหามาตราที่เกี่ยวข้องได้อย่างไร้ปัญหาและเป็นประโยชน์

*ตาราง 8 แสดงผลการทดลองในรูปแบบของ f1-score ของการจำแนกข้อมูลสองประเภทโดยเปรียบเทียบระหว่างแบบจำลองและระหว่างลักษณะข้อมูลรับเข้า*
*โดยกำหนดให้ P หมายถึงคำฟ้อง และ D หมายถึงคำให้การ*

| ประเภทโมเดล | มาตรา 55 | | มาตรา 425 | | มาตรา 438 | | มาตรา 172 | |
|---|---|---|---|---|---|---|---|---|
| | P | P+D | P | P+D | P | P+D | P | P+D |
| MaxEnt (unigram) | 0.235 | 0.179 | 0.247 | 0.359 | **0.232** | **0.205** | **0.164** | **0.263** |
| MaxEnt (unigram+bigram) | 0.215 | 0.157 | 0.319 | 0.290 | 0.206 | 0.152 | 0.130 | 0.222 |
| MaxEnt (TF-IDF) | 0.240 | 0.204 | 0.366 | 0.353 | 0.143 | 0.073 | 0.085 | 0.250 |
| Bi-LSTM | **0.348** | 0.267 | **0.521** | 0.384 | 0.107 | 0.200 | 0.097 | 0.209 |
| CNN | 0.310 | **0.287** | 0.511 | **0.447** | 0.202 | 0.143 | **0.164** | 0.124 |

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากการทดลองในตารางที่ 8 พบว่าการใช้คำฟ้องร่วมกับคำให้การเป็นข้อมูลรับเข้าสำหรับแบบจำลองไม่ได้ทำให้ประสิทธิภาพในการทำนายมาตราที่เกี่ยวข้องดีขึ้นเมื่อเปรียบเทียบกับการใช้คำฟ้องเป็นข้อมูลรับเข้าเพียงอย่างเดียว แต่พบว่าในบางกฎหมาย เช่น ประมวลกฎหมายวิธีพิจารณาความแพ่ง มาตรา 172 การใช้คำให้การร่วมกับคำฟ้องส่งผลให้ประสิทธิภาพของแบบจำลองดีขึ้นในแบบจำลองส่วนใหญ่ สาเหตุอาจเกิดจากมาตรานี้เป็นมาตราที่ใช้สำหรับการโต้แย้งของจำเลย ค่าสัญญาณที่บ่งบอกถึงมาตรานี้จึงมีแนวโน้มที่จะอยู่ในคำให้การด้วย ดังนั้นคำให้การจึงมีส่วนช่วยอย่างมากกับแบบจำลองที่ทำนายมาตราลักษณะนี้

นอกจากนี้พบว่าแบบจำลอง Bi-LSTM และแบบจำลอง CNN มีประสิทธิภาพดีกว่าแบบจำลอง MaxEnt ในประมวลกฎหมายวิธีพิจารณาความแพ่ง มาตรา 55 และประมวลกฎหมายแพ่งและพาณิชย์ มาตรา 425 สาเหตุอาจเกิดจากประมวลกฎหมายวิธีพิจารณาความแพ่ง มาตรา 55 เป็นมาตราที่ปรากฏค่อนข้างบ่อยในคำพิพากษาจำนวนมาก การใส่คำที่บ่งบอกมาตรานี้ในคำพิพากษาอย่างชัดเจนนั้นจึงอาจขึ้นอยู่กับดุลยพินิจของผู้เขียน ส่งผลให้เนื้อหามีความซับซ้อนและยากต่อการพิจารณาเฉพาะในระดับคำเพียงอย่างเดียว อีกทั้ง แบบจำลอง Bi-LSTM และ CNN มีประสิทธิภาพดีกว่าในการทำนายระดับประโยค และสามารถจับค่าสัญญาณจากกลุ่มคำหลายรูปแบบ ตามลำดับ ผลลัพธ์ของมาตราเหล่านี้จึงออกมาดีกว่าแบบจำลอง MaxEnt ที่จำกัดเฉพาะค่าสัญญาณระดับคำ ในขณะเดียวกัน คณะผู้วิจัยสังเกตได้ว่าคำศัพท์ที่ใช้ในประมวลกฎหมายแพ่งและพาณิชย์ มาตรา 425 ซึ่งมีเนื้อหาเกี่ยวข้องกับนายจ้างและลูกจ้างมีลักษณะค่อนข้างเฉพาะ ผลของแบบจำลอง MaxEnt ในมาตรานี้จึงดีกว่าในมาตราอื่น ๆ เป็นพิเศษ อย่างไรก็ตามประสิทธิภาพของแบบจำลองยังคงต่ำกว่าแบบจำลอง Bi-LSTM และแบบจำลอง CNN

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

*ตาราง 9 ผลการเปรียบเทียบค่า f1-score ของแต่ละสถาปัตยกรรม*

| **ประเภทสถาปัตยกรรม** | **ประเภทโมเดล** | **Few Shot** | | | | **Zero Shot** |
|---|---|---|---|---|---|---|
| | | **มาตรา 55** | **มาตรา 425** | **มาตรา 438** | **มาตรา 172** | **มาตราทั้งหมดที่ไม่เกี่ยวกับมาตรา 420** |
| Siamese Neural Network | MaxEnt | 0.2000 | **0.4416** | 0.2025 | 0.1565 | 0.0169 |
| | Bi-LSTM | **0.3162** | 0.2541 | **0.2643** | **0.1871** | 0.1049 |
| | CNN | **0.3162** | 0.2406 | 0.1168 | 0.0556 | 0.0262 |
| | BERT (Cosine | 0.2652 | 0.0916 | 0.1395 | 0.1223 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | similarity) | | | | | |
| Binary Classification | MaxEnt (unigram) | 0.2352 | 0.2469 | 0.2315 | 0.1639 | |
| | MaxEnt (uni+bigram) | 0.2150 | 0.3188 | 0.2058 | 0.1304 | |
| | MaxEnt (TfIdf) | 0.2407 | 0.3661 | 0.1428 | 0.0851 | |
| | Bi-LSTM | 0.3478 | 0.5210 | 0.1684 | 0.2000 | |
| | CNN | 0.3097 | 0.5106 | 0.2016 | 0 .1643 | |
| | BERT (NSP) | 0.1446 | 0.0604 | 0.0455 | 0.0952 | |

จากผลการทดลองในตารางที่ 9 พบว่าแบบจำลอง Bi-LSTM ทำนายได้ดีที่สุดในมาตราส่วนใหญ่ แต่ไม่ได้มีประสิทธิภาพที่ดีกว่าอย่างเห็นได้ชัด

นอกจากนี้ คณะผู้วิจัยพบว่าประสิทธิภาพการทำนายมาตราที่เกี่ยวข้องของแบบจำลองบนโครงข่ายประสาทเทียมแบบสยามมีความใกล้เคียงกับแบบจำลองการจำแนกข้อมูลสองประเภทในประมวลกฎหมายแพ่งและพาณิชย์มาตรา 438 และประมวลกฎหมายวิธีพิจารณาความแพ่งมาตรา 172 ของแบบจำลอง Bi-LSTM แม้ว่าแบบจำลองบนโครงข่ายประสาทเทียมสยามจะไม่ได้เรียนรู้คำฟ้องของแต่ละมาตรามาโดยเฉพาะ เช่นเดียวกับแบบจำลองการจำแนกข้อมูลสอบประเภท จากผลการทดลองนี้อาจสรุปได้ว่าแบบจำลองโครงข่ายประสาทเทียมแบบสยามมีความเป็นไปได้ที่จะสามารถทำนายมาตราที่แบบจำลองไม่ได้เรียนรู้จากข้อมูลชุดฝึก

ในขณะที่แบบจำลองที่ปราศจากข้อมูลปฐมภูมิในการเรียนรู้ (Zero-shot) ทำนายได้ดีน้อยกว่าแบบจำลองที่พัฒนาขึ้นด้วยข้อมูลชุดฝึกที่มีจำนวนน้อย (Few-shot) อย่างไรก็ตาม แม้แบบจำลองจะไม่มีข้อมูลในการเรียนรู้ประกอบ แต่ประสิทธิภาพการทำนายมาตราที่เกี่ยวข้องของแบบจำลองนั้นยังสามารถแสดงผลออกมาได้ดีในระดับหนึ่ง ทั้งนี้ อาจมีสาเหตุมาจากแบบจำลองเหล่านี้มีองค์ความรู้ทั่วไปเกี่ยวกับภาษาจากกระบวนการถ่ายทอดการเรียนรู้ (Transfer Learning) ที่แบบจำลองได้รับผ่านเวกเตอร์ของคำที่ได้จากการเรียนรู้ของแบบจำลองอื่นมาก่อน (Pretrained Word Embedding) แล้วในระดับหนึ่ง

สำหรับประเภทแบบจำลองตัวเข้ารหัสแบบสองทิศทางในรูปแบบตัวแปลง (BERT) มีประสิทธิภาพการทำนายที่ต่ำที่สุดในทั้งสองสถาปัตยกรรม อาจมีสาเหตุมาจากความยาวเอกสารทางกฎหมายมีความยาวมากเกินกว่าที่แบบจำลองจะเก็บข้อมูลได้หมด และแบบจำลอง BERT นั้น

ถูกฝึกมาจากฐานข้อมูลขององค์ความรู้ทั่วไป การนำมาใช้กับเอกสารทางกฎหมายที่ปรากฏการใช้คำเฉพาะทาง จึงอาจทำให้แบบจำลองไม่สามารถเก็บค่าสัญญาณเหล่านี้ได้ครบถ้วน

*ตาราง 10 ผล f1-score@3 ของโครงข่ายประสาทเทียมสยามโดยเปรียบเทียบผลระหว่างแบบจำลอง*

| ประเภทโมเดล | Few-shot | | | | Zero-shot |
|---|---|---|---|---|---|
| | มาตรา 55 | มาตรา 425 | มาตรา 438 | มาตรา 172 | มาตราทั้งหมดที่ไม่เกี่ยวกับมาตรา 420 |
| Bi-LSTM | 0.5057 | 0.4083 | 0.4091 | 0.3086 | 0.3164 |
| CNN | 0.5678 | 0.4062 | 0.4072 | 0.2737 | 0.3307 |
| BERT | 0.6398 | 0.3137 | 0.1212 | 0.1089 | |
| Elasticsearch | 0.3759 | 0.0000 | 0.3853 | 0.2021 | |

จากผลการทดลองในตารางที่ 10 ซึ่งเป็นการทดลองในการค้นคืนมาตราที่เกี่ยวข้องโดยใช้แบบจำลองบนโครงข่ายประสาทเทียมสยาม และวัดผลด้วยค่า f1-score@3 พบว่าแบบจำลองมีประสิทธิภาพที่ดีขึ้นอย่างมากในทุกมาตราอย่างสม่ำเสมอ ไม่ว่าจะเป็นแบบจำลองที่ฝึกด้วยข้อมูลชุดฝึกจำนวนน้อย (Few-shot) หรือแบบจำลองที่ไม่มีข้อมูลชุดฝึกเลย (Zero-shot) นอกจากนี้ เมื่อเปรียบเทียบกับผลการค้นหามาตราที่เกี่ยวข้องด้วยเครื่องมือค้นหาและสืบค้นข้อมูล (Elasticsearch) แบบจำลองบนโครงข่ายประสาทเทียมสยามที่คณะผู้วิจัยพัฒนาขึ้นก็มีประสิทธิภาพดีกว่าเช่นกัน ซึ่งแสดงให้เห็นว่าแบบจำลองมีความสามารถในการคัดกรองและเลือกมาตราที่เกี่ยวข้องต่อคำพิพากษาได้ดีได้ดี

# บทที่ 5
# สรุปผล

จากการวิจัยนี้ คณะผู้วิจัยได้รวบรวมข้อมูลคำพิพากษาทั้งหมด 80,654 คำพิพากษาตั้งแต่ปี พ.ศ. 2471 ถึงปีพ.ศ. 2559 และแบ่งข้อมูลเหล่านี้ออกเป็นสามกลุ่มหลัก คือ คำฟ้อง คำให้การ และป้ายกำกับมาตราที่เกี่ยวข้อง โดยมีจำนวนคำพิพากษาทั้งหมดอยู่ที่ 4,448 คำพิพากษา นอกจากนี้ คณะผู้วิจัยยังรวบรวมข้อมูลกฎหมายแพ่งและพาณิชย์ที่เกี่ยวข้องกับคำพิพากษา เพื่อใช้เป็นค่าสัญญาณหลักในการเรียนรู้ของแบบจำลองต่างๆ ในการคัดกรองและเลือกมาตราที่เกี่ยวข้องกับข้อเท็จจริงในคำพิพากษา เพื่อพัฒนาระบบค้นคืนมาตราที่เกี่ยวข้อง

ในการพัฒนาแบบจำลอง คณะผู้วิจัยใช้แบบจำลอง Bi-LSTM, CNN, และ BERT บนสถาปัตยกรรมของการจำแนกข้อมูลสองประเภท เพื่อทดสอบการใช้คำให้การเป็นข้อมูลรับเข้าร่วมกับคำฟ้อง จากผลการทดลองพบว่าการเพิ่มคำให้การไม่มีผลอย่างชัดเจนต่อประสิทธิภาพในการทำนายมาตราที่เกี่ยวข้องของแบบจำลองส่วนใหญ่ นอกจากนี้ แบบจำลองโครงข่ายประสาทเทียมสยามยังมีประสิทธิภาพการทำนายดีกว่าเครื่องมือค้นหาและสืบค้นข้อมูล (Elasticsearch) ซึ่งเป็นตัวชี้วัดในการสร้างระบบค้นคืน และมีคะแนน f1-score@3 มากกว่าร้อยละ 50 ในมาตรา 55 จากผลการทดลองทั้งหมด จากผลลัพธ์ที่ได้ สรุปได้ว่าแบบจำลองที่พัฒนาขึ้นอาจนำไปพัฒนาเป็นระบบค้นคืนมาตราที่ครอบคลุมข้อกฎหมายอย่างกว้างขึ้น และมีประโยชน์ในการนำไปประยุกต์ใช้กับคำพิพากษาในบริบทที่แตกต่างออกไป โดยไม่จำเป็นต้องเริ่มต้นฝึกแบบจำลองใหม่ ซึ่งเป็นการประหยัดเวลาและทรัพยากร และแก้ไขปัญหาข้อจำกัดของปริมาณข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

# บรรณานุกรม


Alammar, J. (2018). "The Illustrated Transformer." Retrieved 01032023, 2023, from https://jalammar.github.io/illustrated-transformer/.

Alarie, B. and K. Gardhouse (2021). "Predicting Worker Classification in the Gig Economy." Tax Notes Federal: 1733.

Aletras, N., et al. (2016). "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective." PeerJ Computer Science **2**: e93.

Chalkidis, I., et al. (2019). "Neural legal judgment prediction in english." arXiv preprint arXiv:1906.02059.

Chormai, P., et al. (2019). "AttaCut: A Fast and Accurate Neural Thai Word Segmenter." arXiv preprint arXiv:1911.07056.

Dunn, M., et al. (2017). Early predictability of asylum court decisions. Proceedings of the 16th edition of the International Conference on Articial Intelligence and Law.

Ismail, A. A., et al. (2018). "Improving long-horizon forecasts with expectation-biased LSTM networks." arXiv preprint arXiv:1804.06776.

Iyyer, M., et al. (2015). Deep unordered composition rivals syntactic methods for text classification. Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers).



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Jurafsky, D. and J. H. Martin (2008). "Speech and Language Processing: An introduction to speech recognition, computational linguistics and natural language processing." Upper Saddle River, NJ: Prentice Hall.

Katz, D. M., et al. (2017). "A general approach for predicting the behavior of the Supreme Court of the United States." PloS one **12**(4): e0174698.

Koch, G., et al. (2015). Siamese neural networks for one-shot image recognition. ICML deep learning workshop, Lille.

Kowsrihawat, K., et al. (2018). Predicting judicial decisions of criminal cases from thai supreme court using bi-directional GRU with attention mechanism. 2018 5th Asian Conference on Defense Technology (ACDT), IEEE.

LeCun, Y., et al. (2015). "Deep learning." nature **521**(7553): 436-444.

Thammaboosadee, S. and U. Silparcha (2008). A framework for criminal judicial reasoning system using data mining techniques. 2008 2nd IEEE International Conference on Digital Ecosystems and Technologies, IEEE.

Thammaboosadee, S., et al. (2014). "A Two-Stage Classifier That Identifies Charge and Punishment under Criminal Law of Civil Law System." IEICE TRANSACTIONS on Information and Systems **97**(4): 864-875.

Zhang, Y. and B. Wallace (2015). "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification." arXiv preprint arXiv:1510.03820.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

# ประวัติผู้เขียน

| | |
|---|---|
| **ชื่อ-สกุล** | Thaweesak Chusri |
| **วัน เดือน ปี เกิด** | 01 Jan 1991 |
| **สถานที่เกิด** | Nakhon Si Thammarat |
| **วุฒิการศึกษา** | Chulalongkorn University |
| **ที่อยู่ปัจจุบัน** | 46/359 Life Asoke, Asoke-Dindaeng, Bang Kapi, Huai Khwang, Bangkok 10310 |



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY