แบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญาที่เรียนรู้จากคำพิพากษาศาลฎีกาไทย โดยใช้เทคนิคการเรียนรู้เชิงลึก

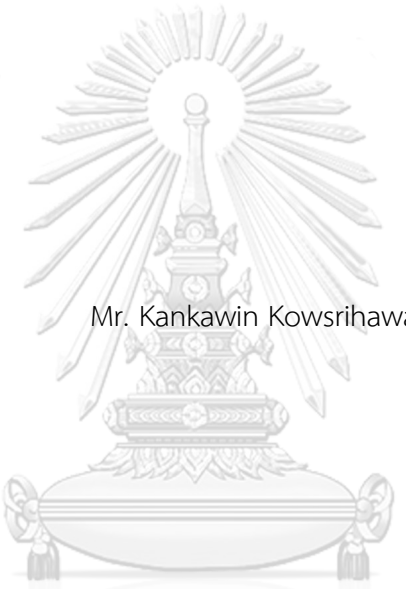
นายกานกวิญจน์ โค้วสีหวัฒน์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A Criminal Case Outcome and Issue Prediction Model On Thai Supreme Court Cases Using Deep Learning Techniques

Mr. Kankawin Kowsrihawat



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2018

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	แบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญาที่เรียนรู้จากคำพิพากษาศาลฎีกาไทย โดยใช้เทคนิคการเรียนรู้เชิงลึก
โดย	นายกานกวิญจน์ โค้วสีหวัฒน์
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.พีรพล เวทีกูล
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	อาจารย์ ดร.ปรัชญา บุญขวัญ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.พีรพล เวทีกูล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(อาจารย์ ดร.ปรัชญา บุญขวัญ)

..... กรรมการ
(อาจารย์ ดร.เอกพล ช่วงสุวนิช)

..... กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.นวลวรรณ สุนทรภิษัช)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

กานกวิญจน์ โค้วสีหวัฒน์ : แบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญาที่เรียนรู้จากคำพิพากษาศาลฎีกาไทย โดยใช้เทคนิคการเรียนรู้เชิงลึก. (A Criminal Case Outcome and Issue Prediction Model On Thai Supreme Court Cases Using Deep Learning Techniques) อ.ที่ปรึกษาหลัก : ผศ. ดร.พีรพล เวทีกูล, อ.ที่ปรึกษาร่วม : อ. ดร.ปรัชญา บุญขวัญ

การทำนายผลคำตัดสินในคดีอาญาด้วยแบบจำลองการเรียนรู้ของเครื่องได้รับความสนใจมากขึ้นในช่วงเวลาที่ผ่านมา อย่างไรก็ดี เทคนิคที่ใช้ในแบบจำลองดังกล่าวมักใช้ตัวแทนข้อความที่มีที่มาจากแบบจำลองถุงคำ ซึ่งไม่สนใจลำดับของข้อความทำให้สูญเสียบริบทของข้อความ และผลลัพธ์การทำนายมีความแม่นยำลดลง ดังนั้น วิทยานิพนธ์ฉบับนี้จึงเสนอแบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญาซึ่งเรียนรู้จากคำพิพากษาศาลฎีกาไทยโดยใช้เทคนิคการเรียนรู้เชิงลึกผ่านชุดโครงข่ายประสาทเทียม แบบจำลองนี้สร้างตัวแทนข้อความด้วยโครงข่ายประตูวกกลับสองทิศทางร่วมด้วยกลไกจุดสนใจ ก่อนนำตัวแทนข้อความนั้นไปทำนายผลคำตัดสินและประเด็นในคดีอาญาด้วยโครงข่ายประสาทเทียมแบบโมดูลซึ่งจำลองโครงสร้างความรับผิดทางอาญาตามทฤษฎีกฎหมายอาญา

ผลการทดลองแสดงให้เห็นว่าแบบจำลองให้ประสิทธิภาพสูงกว่าแบบจำลองที่ใช้เทคนิคการเรียนรู้ของเครื่องเดิมอย่างเนอีฟเบยส์และเอสวีเอ็ม เมื่อพิจารณาจากค่า F1 นอกจากนี้แบบจำลองยังให้ประสิทธิภาพสูงในการทำนายประเด็นในคดีอาญาบางประเด็นซึ่งมีผลต่อการทำนายผลคำตัดสินในคดีอาญาด้วย นอกจากนั้น ผลการทดลองสะท้อนให้เห็นว่า การใช้โครงข่ายประตูวกกลับสองทิศทางร่วมด้วยกลไกจุดสนใจสามารถสร้างตัวแทนข้อความที่ดีกว่าแบบจำลองดั้งเดิมที่มีลักษณะเดียวกันกับถุงคำ ตลอดจนโครงข่ายประสาทเทียมแบบโมดูลสามารถจำลองโครงสร้างความรับผิดทางอาญาได้

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่อนิสิต .....
ปีการศึกษา 2561 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

5970903421 : MAJOR COMPUTER SCIENCE

KEYWORD: Legal case prediction, Case outcome and issue prediction, Criminal case, Thai Supreme Court cases, Deep learning

Kankawin Kowsrihawat : A Criminal Case Outcome and Issue Prediction Model On Thai Supreme Court Cases Using Deep Learning Techniques. Advisor: Asst. Prof. Peerapon Vateekul, Ph.D. Co-advisor: Prachya Boonkwan, Ph.D.

Predicting court judgement using Machine learning models has been interesting over the past years. However, traditional techniques used the text representation based on Bag-of-Words model (BoW), where the order of words is discarded, resulting in context loss and low accuracy. In this thesis, we propose a prediction model of criminal cases from Thai Supreme Court using end-to-end deep learning neural networks. This model creates the text representation using the bi-directional version of Gated Recurrent Unit (GRU) with Attention mechanism. Then, it predicts the case outcome and legal issues from that case using the modular neural network which follows the logic of the sophisticated criminal law structure.

After the experiment, in the case outcome prediction, this model yields higher F1 scores than traditional text classification techniques including Naïve Bayes and SVM. For some legal issues prediction related to the case outcome, this model yields higher F1 scores than original techniques. Moreover, the experiment shows that the text representation from the bi-directional GRU with Attention mechanism is better than the BoW-based model. Besides, the modular network is shown to represent the criminal law structure.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

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

Academic Year: 2018 Advisor's Signature .....

Co-advisor's Signature .....

กิตติกรรมประกาศ

การทำวิทยานิพนธ์ฉบับนี้เป็นความท้าทายประการหนึ่งในชีวิตของผู้วิจัย ด้วยความมุ่งหวังที่จะประยุกต์ใช้ความรู้ทางวิทยาศาสตร์คอมพิวเตอร์ร่วมกับความรู้ทางกฎหมาย นอกจากนี้ ตลอดระยะเวลาการวิจัยและทำวิทยานิพนธ์ ผู้วิจัยเผชิญอุปสรรคและปัญหามากมาย ทั้งความเข้าใจในแนวคิดและทฤษฎีที่เกี่ยวข้อง การทำการทดลองที่ถูกต้องสอดคล้องกับวิธีวิทยาการวิจัย ตลอดจนภาระหน้าที่การงานของผู้วิจัยในฐานะบุคคลที่ทำงานไปด้วยและเรียนไปด้วย ดังนั้น เมื่อวิทยานิพนธ์นี้จัดทำขึ้นได้เป็นผลสำเร็จ ผู้วิจัยต้องขอขอบพระคุณบุคคลที่มีส่วนสำคัญต่อความสำเร็จของวิทยานิพนธ์ฉบับนี้

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.พีรพล เวทีกูล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งกรุณารับเป็นที่ปรึกษาต่อเนื่องจากที่ผู้วิจัยเคยเป็นศิษย์ในที่ปรึกษาโครงการทางวิศวกรรมคอมพิวเตอร์ในระดับชั้นปริญญาตรี ท่านพยายามอย่างเต็มที่ในการให้ความช่วยเหลือ ให้กำลังใจ และให้คำแนะนำต่าง ๆ โดยเฉพาะกระบวนการทางวิศวกรรมและวิทยาศาสตร์ข้อมูล อันเป็นองค์ความรู้ที่จะติดตัวผู้วิจัยไปตลอดชีวิต ("สถิติข้อมูลสำคัญมาก")

ขอขอบพระคุณ อาจารย์ ดร.ปรัชญา บุญขวัญ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ซึ่งเคยเป็นผู้ให้คำแนะนำในการทำโครงการทางวิศวกรรมคอมพิวเตอร์ในระดับชั้นปริญญาตรีของผู้วิจัยมาแล้ว ท่านอาจารย์ให้ความเมตตากรุณาต่อผู้วิจัยอย่างมาก ทั้งในแง่การให้ความรู้ทางการวิจัย องค์ความรู้ด้านการประมวลผลภาษาธรรมชาติร่วมกับเทคนิคการเรียนรู้เชิงลึก ตลอดจนคอยสละเวลาอันมีค่าของท่านในการให้คำแนะนำต่าง ๆ เป็นอย่างดี

ขอขอบพระคุณ ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล ประธานกรรมการสอบวิทยานิพนธ์ ท่านอาจารย์ได้กรุณาให้ข้อสังเกตกับคำแนะนำที่เฉียบคมและมีค่าอย่างยิ่งต่อผู้วิจัย เพื่อให้องค์ความรู้ทางวิทยาศาสตร์คอมพิวเตอร์และกฎหมายสามารถประสานรวมกันกลายเป็นวิทยานิพนธ์ฉบับนี้ได้สำเร็จ

ขอขอบพระคุณ อาจารย์ ดร.เอกพล ช่วงสุวนิช กรรมการสอบวิทยานิพนธ์ ท่านอาจารย์ได้กรุณาให้ข้อสังเกตกับคำแนะนำด้านการประมวลผลภาษาธรรมชาติด้วยเทคนิคการเรียนรู้เชิงลึกซึ่งมีคุณค่าต่อวิทยานิพนธ์ฉบับนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ขอขอบพระคุณ รองศาสตราจารย์ ดร.นวลวรรณ สุนทรภิษัช กรรมการผู้ทรงคุณวุฒิจากภายนอกมหาวิทยาลัย ซึ่งท่านอาจารย์กรุณาให้ข้อสังเกตและคำแนะนำที่ดียิ่งในฐานะเป็นผู้เชี่ยวชาญในการประยุกต์องค์ความรู้ด้านวิทยาศาสตร์คอมพิวเตอร์เข้ากับความรู้ด้านกฎหมาย ช่วยให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณ คณาจารย์ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย คณาจารย์คณะนิติศาสตร์ มหาวิทยาลัยรามคำแหง และคณาจารย์สำนักอบรมศึกษากฎหมายแห่งเนติบัณฑิตยสภา ซึ่งอบรมสั่งสอนองค์ความรู้ต่าง ๆ ต่อผู้วิจัย และเป็นแรงบันดาลใจที่ทำให้วิทยานิพนธ์ฉบับนี้เกิดขึ้นมาได้

ขอขอบคุณ "พี่แคน" คุณแคน อุดมเจริญชัยกิจ นิสิตปริญญาเอกภาควิชาวิศวกรรมคอมพิวเตอร์ และ "พี่บอส" คุณณัฐชัย ตรีทศายุธ วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ ซึ่งทั้งสองท่านเป็นรุ่นพี่ในห้องปฏิบัติการที่คอยให้คำแนะนำและความช่วยเหลือต่าง ๆ ต่อผู้วิจัย ช่วยให้ผู้ซึ่งไม่มีองค์ความรู้ด้านเทคนิคการเรียนรู้เชิงลึกมาก่อนเลยเช่นผู้วิจัย ได้ซึมซับและเข้าใจองค์ความรู้นี้มากขึ้นทั้งทางทฤษฎีและเทคนิค จนนำไปสู่ความสำเร็จของวิทยานิพนธ์ฉบับนี้ นอกจากนี้ขอขอบคุณพี่ ๆ เพื่อน ๆ และน้อง ๆ ในห้องปฏิบัติการท่านอื่น ๆ ที่ช่วยให้การสนับสนุนต่าง ๆ ในการทำวิทยานิพนธ์ฉบับนี้

ท้ายที่สุด ขอกราบขอบพระคุณ คุณหญิงประณีตศิลป์ วัชรพล ผู้เมตตาส่งเสียให้ผู้วิจัยได้มีโอกาสศึกษาต่อในระดับปริญญาโทและมีโอกาสจัดทำวิทยานิพนธ์ฉบับนี้ ตลอดจนขอขอบคุณครอบครัว โดยเฉพาะอย่างยิ่ง คุณแม่นิสาสินี โค้วสีหวัฒน์ ที่คอยเลี้ยงดูสั่งสอนและฟูมฟักให้ "อภิชาตบุตร" ที่ท่านเอ่ยถึงทุกครั้งเดินทางมาถึงจุดนี้ได้

อนึ่ง วิทยานิพนธ์ฉบับนี้ให้บทเรียนชีวิตแก่ผู้วิจัยหลายอย่าง ที่สำคัญคือ การเรียนรู้ที่จะปรับปรุงตนเองจากความผิดพลาดเหมือนการปรับค่าน้ำหนักของโครงข่ายประสาทเทียม ตลอดจนความสำคัญของเวลาที่แม้เพียงมิลลิวินาทีเดียวก็มีค่าเสมอ

กานกวิญจน์ โค้วสีหวัฒน์

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ค

บทคัดย่อภาษาอังกฤษ ..... ง

กิตติกรรมประกาศ ..... จ

สารบัญ ..... ช

สารบัญตาราง ..... ฎ

สารบัญรูปภาพ ..... ฐ

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญ 1

1.2 วัตถุประสงค์ของการวิจัย 3

1.3 ขอบเขตของการวิจัย 3

1.4 ประโยชน์ที่คาดว่าจะได้รับ 4

1.5 วิธีดำเนินการวิจัย 4

1.6 ผลงานวิจัยที่ตีพิมพ์ 5

บทที่ 2 แนวคิดและทฤษฎีเกี่ยวข้อง 7

2.1 ศาลฎีกาและคำพิพากษาศาลฎีกา 7

2.2 ความรู้เบื้องต้นเกี่ยวกับกฎหมายอาญา 8

2.3 การวินิจฉัยคดีอาญาด้วยโครงสร้างความรับผิดทางอาญา 9

2.3.1 การพยายามกระทำความผิด 12

2.4 ตัวแทนข้อความ (Text representation) 14

2.5 โครงข่ายประสาทเทียม (Artificial Neural Network : ANN) 18

2.5.1 เพอร์เซปตรอน (Perceptron) [14] 18

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.5.2 โครงข่ายประสาทเทียมแบบไปข้างหน้า (Feedforward neural network)..... 19
2.5.3 กฎการสอนโครงข่ายประสาทเทียมแบบไปข้างหน้า..... 21
2.5.4 การหยุดเรียนรู้กลางคัน (Drop-out) [17]..... 24
2.6 การเรียนรู้เชิงลึก (Deep learning) [19]..... 25
2.7 โครงข่ายประสาทเทียมแบบวกกลับ (Recurrent Neural Network : RNN) [19]..... 25
2.7.1 ปัญหาการพึ่งพาระยะยาว (Long-term dependency)..... 26
2.7.2 โครงข่ายความจำระยะสั้นขนาดยาว (Long Short-Term Memory : LSTM) [20, 21] 27
2.7.3 โครงข่ายประตูวกกลับ (Gated Recurrent Unit : GRU) [22]..... 28
2.8 กลไกจุดสนใจ (Attention Mechanism)..... 29
2.9 การวัดประสิทธิภาพการทำงาน (Performance evaluation)..... 31
บทที่ 3 งานวิจัยที่เกี่ยวข้อง..... 34
3.1 งานวิจัยเกี่ยวกับการทำนายผลคำพิพากษาของศาลในต่างประเทศ..... 34
3.1.1 ผลงานวิจัยโดย Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preotiuc-Pietro และ Vasileios Lampos (ค.ศ.2016) [27]..... 34
3.1.2 ผลงานวิจัยโดย Daniel Martin Katz, Michael J. Bommarito II และ Josh Blackman (ค.ศ.2017) [28]..... 36
3.2 งานวิจัยทางคอมพิวเตอร์เกี่ยวกับคำพิพากษาศาลฎีกาไทย..... 38
3.2.1 ผลงานวิจัยโดย โษฑศ์รัตต ธรรมบุษดี และ อุดม ศิลปอาชา (ค.ศ.2008) [30]..... 38
3.2.2 ผลงานวิจัยโดย โษฑศ์รัตต ธรรมบุษดี, บัณฑิต วรรธนาภา, Jonathan H. Chan และ อุดม ศิลปอาชา (ค.ศ.2014) [31]..... 39
3.2.3 ผลงานวิจัยโดย พงศ์ปณัฏฐ โอสถิตย์พร, นวลวรรณ สุนทรภิษัช และ วิวัฒน์ วัฒนาวุฒิ (ค.ศ.2017) [32]..... 42
3.2.4 ผลงานวิจัยโดย ธนพล ตันติศรีปรีชา และ นวลวรรณ สุนทรภิษัช (ค.ศ.2016) [33]... 42
3.2.5 ผลงานวิจัยโดย กานกวิญจน์ โค้วสีหวัฒน์ และ พีรพล เวทีกูล (ค.ศ.2015) [8]..... 43

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 4 แนวคิดและวิธีการดำเนินงาน 44
4.1 แบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญาโดยใช้เทคนิคการเรียนรู้เชิงลึก 44
4.2 การประมวลผลข้อมูลเบื้องต้น (Preprocessing) 45
4.3 กระบวนการทำงานภายในแบบจำลอง 46
4.3.1 การฝังคำ (Word embedding) 46
4.3.2 การสร้างตัวแทนข้อความด้วยโครงข่ายประตูวกกลับสองทิศทาง (Bi-directional GRU : Bi-GRU) 46
4.3.3 การเพิ่มประสิทธิภาพให้ตัวแทนข้อความด้วยกลไกจุดสนใจ 48
4.3.4 การทำนายผลคำตัดสินและประเด็นในคดีอาญาด้วยโครงข่ายประสาทเทียมแบบโมดูล 48
4.4 ชุดข้อมูลที่ใช้ในการทดลอง 51
4.4.1 การสร้างตารางข้อมูลคำพิพากษา 53
4.4.2 การสร้างตารางข้อมูลข้อกฎหมาย 56
บทที่ 5 การทดลองและผลการทดลอง 58
5.1 สมมติฐานการทดลอง 58
5.2 การออกแบบการทดลอง 58
5.2.1 การแบ่งชุดข้อมูล 58
5.2.2 แบบจำลองที่นำเสนอซึ่งใช้ในการทดลอง 59
5.2.3 แบบจำลองตั้งต้น (Baseline model) 61
5.3 ผลการทดลอง 63
5.3.1 การทำนายผลคำตัดสิน 63
5.3.2 การทำนายประเด็นในคดีอาญา 67
บทที่ 6 สรุปผลการทดลองและแนวทางการวิจัยในขั้นถัดไป 74
6.1 สรุปผลการทดลอง 74

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

6.2 แนวทางการวิจัยในขั้นถัดไป 75
ภาคผนวก ก. ตัวอย่างแถวข้อมูลในตารางต่าง ๆ ของชุดข้อมูลทีเอสซีซี 79
บรรณานุกรม 83
ประวัติผู้เขียน 85

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญตาราง

หน้า

ตารางที่ 1.1 ตารางแสดงแผนปฏิบัติงานตามวิธีดำเนินการวิจัย..... 6

ตารางที่ 2.1 ตารางแสดงประพจน์ที่พิจารณาได้จากโครงสร้างความรับผิดทางอาญา 10

ตารางที่ 4.1 ตารางแสดงที่มาและจำนวนของคำในพจนานุกรมที่ใช้ในซอฟต์แวร์ตัดคำ ไพเล็กซ์โต สำหรับแบบจำลองของผู้วิจัย 45

ตารางที่ 4.2 ตารางแสดงหมวดความผิดและฐานความผิดตามประมวลกฎหมายอาญาของคำพิพากษาทั้งหมดที่นำมาใช้เป็นข้อมูลในฐานข้อมูลทีเอสซีซี 51

ตารางที่ 4.3 ตารางแสดงคุณลักษณะของแถวข้อมูลหนึ่งในตาราง คำพิพากษา ในชุดข้อมูลทีเอสซีซี 53

ตารางที่ 4.4 ตารางแสดงสถิติข้อมูลของตาราง คำพิพากษา ในชุดข้อมูลทีเอสซีซี 55

ตารางที่ 4.5 ตารางแสดงสถิติข้อมูลส่วนประเด็นในคดีอาญาของตาราง คำพิพากษา ในชุดข้อมูลทีเอสซีซี 56

ตารางที่ 4.6 ตารางแสดงคุณลักษณะของแถวข้อมูลหนึ่งในตาราง กฎหมาย ในชุดข้อมูลทีเอสซีซี 56

ตารางที่ 4.7 ตารางแสดงสถิติข้อมูลของตาราง ข้อกฎหมาย ในชุดข้อมูลทีเอสซีซี 57

ตารางที่ 5.1 ตารางแสดงสถิติชุดข้อมูลที่แบ่งออกจากชุดข้อมูลทีเอสซีซีเพื่อการทดลองทดสอบประสิทธิภาพของแบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญา..... 59

ตารางที่ 5.2 ตารางแสดงพารามิเตอร์ที่ทดลองปรับกับแบบจำลอง และชุดพารามิเตอร์ที่ทำให้ แบบจำลองได้ค่าความแม่นยำสูงที่สุดในชุดข้อมูลพัฒนา 61

ตารางที่ 5.3 ตารางแสดงพารามิเตอร์ที่ดีที่สุดสำหรับแบบจำลองไม่มีกลไกจุดสนใจ และแบบจำลองจุดสนใจในตัวเองแยกข้อความ ซึ่งใช้เป็นแบบจำลองตั้งต้นในการทดลอง 62

ตารางที่ 5.4 ตารางแสดงผลการทดลองวัดประสิทธิภาพการทำนายผลคำตัดสินในคดีอาญาจาก แบบจำลองต่าง ๆ ผลที่ดีที่สุดจะแสดงเป็นตัวหนาและขีดเส้นใต้..... 63

ตารางที่ 5.5 ตารางจำนวนคำตอบที่แบบจำลองแต่ละแบบทำนายได้เปรียบเทียบกับจำนวนคำตอบที่แบบจำลองตอบได้ถูกต้อง ผลที่ดีที่สุดจะแสดงเป็นตัวหนาและขีดเส้นใต้ 65

ตารางที่ 5.6 ตารางแสดงสถิติข้อมูลส่วนประเด็นในคดีอาญาในชุดข้อมูลสอน 68

ตารางที่ 5.7 ตารางแสดงสถิติข้อมูลส่วนประเด็นในคดีอาญาในชุดข้อมูลทดสอบ 68

ตารางที่ 5.8 ตารางแสดงผลการทดลองวัดประสิทธิภาพการทำนายประเด็นในคดีอาญาจากแบบจำลองต่าง ๆ โดยถือว่าประเด็นที่มีผลลัพธ์ -1 มีค่าเป็น 0 ผลที่ดีที่สุดจะแสดงเป็นตัวหนาและขีดเส้นใต้ 70

ตารางที่ 5.9 ตารางแสดงผลการทดลองวัดประสิทธิภาพการทำนายประเด็นในคดีอาญาจากแบบจำลองต่าง ๆ โดยตัดประเด็นที่มีผลลัพธ์เป็น -1 ออกไป ผลที่ดีที่สุดจะแสดงเป็นตัวหนาและขีดเส้นใต้ 71

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญรูปภาพ

หน้า

รูปที่ 1.1 กระบวนการตัดสินคดีอาญาของผู้พิพากษาโดยใช้โครงสร้างความรับผิดทางอาญา 1

รูปที่ 1.2 โครงสร้างของแบบจำลองทำนายผลคำตัดสินคดีอาญาที่ผู้วิจัยพัฒนาขึ้น 2

รูปที่ 2.1 ลอจิกเกตซึ่งเกิดจากกฎทางตรรกะแสดงโครงสร้างความรับผิดทางอาญา 11

รูปที่ 2.2 ตัวอย่างการสื่อความหมายทางภาษาของเวกเตอร์คำด้วยระยะห่างระหว่างเวกเตอร์ [13] 17

รูปที่ 2.3 โครงสร้างของเพอร์เซปตรอน 18

รูปที่ 2.4 โครงสร้างของโครงข่ายประสาทเทียมแบบไปข้างหน้า [16] 19

รูปที่ 2.5 สเปซแสดงค่าความผิดพลาดในการสอนเพอร์เซปตรอนหรือหน่วยย่อย เมื่อกำหนดค่าน้ำหนักของข้อมูลขาเข้า $w1$ และค่าน้ำหนักพิเศษ $w0$ แตกต่างกัน [14] 23

รูปที่ 2.6 ภาพแสดงแนวคิดการหยุดเรียนรู้กลางคัน หน่วยย่อยที่ถูกหยุดเรียนรู้กลางคัน แสดงตามภาพขวาด้วยวงกลมที่มีกากบาทภายใน [17] 24

รูปที่ 2.7 ภาพซ้ายแสดงการหยุดการเรียนรู้กลางคันในช่วงสอนข้อมูลให้กับแบบจำลอง โดยสุ่มหยุดการทำงานของ หน่วยย่อยด้วยค่าความน่าจะเป็น p ส่วนภาพขวาแสดงวิธีการใช้งานแบบจำลองในช่วงทดสอบข้อมูล โดย ไม่มี การหยุดการทำงานของหน่วยย่อยใด แต่จะถ่วงน้ำหนักเส้นทางของโครงข่ายที่จะออกไปหาหน่วยย่อยอื่นแทน [17] 24

รูปที่ 2.8 โครงสร้างของโครงข่ายประสาทเทียมแบบวกกลับ ซึ่งความจริงแล้วคือการนำโครงข่ายประสาทเทียมธรรมดามาต่อกันเพื่อรับข้อมูลที่มีลักษณะเป็นลำดับ หรือมาตามเวลานั่นเอง [19] 26

รูปที่ 2.9 ตัวอย่างการประยุกต์ใช้กลไกจุดสนใจตาม [24] เพื่อแปลข้อความภาษาอังกฤษ “The defendant defames the plaintiff.” เป็นภาษาไทยว่า “จำเลยหมิ่นประมาทโจทก์” ด้วยโครงข่ายประสาทเทียมแบบวกกลับ 29

รูปที่ 4.1 โครงสร้างภายในแบบจำลองทำนายผลคำตัดสินคดีอาญาที่เรียนรู้จาก คำพิพากษาศาลฎีกาไทย โดยใช้เทคนิคการเรียนรู้เชิงลึก 44

รูปที่ 4.2 ลอจิกเกตซึ่งได้รับการดัดแปลงเพื่อใช้ในแบบจำลอง 50

จุฬาลงกรณ์มหาวิทยาลัย Chulalongkorn University

รูปที่ 5.1 วิธีการใช้งานแบบจำลองทำนายผลคำตัดสินคดีอาญาที่เรียนรู้จากคำพิพากษาศาลฎีกาไทย โดยใช้เทคนิคการเรียนรู้เชิงลึก ร่วมกับชุดข้อมูล ทีเอสซีซี ทั้งนี้ แสดงเฉพาะส่วนการทำนายผลคำตัดสินเท่านั้น 60

รูปที่ 5.2 โครงสร้างภายในแบบจำลองจุดสนใจในตัวเองแยกข้อความ 62

รูปที่ 5.3 ภาพแสดงค่าความสนใจของคำต่าง ๆ จากการใช้กลไกจุดสนใจในประเด็นปัญหาทดสอบหนึ่ง ค่าความสนใจของคำใดมากจะมีพื้นหลังเป็นสีเขียว และค่าที่น้อยลงจะแสดงเป็นสีเหลือง สีส้ม และสีแดงตามลำดับ อนึ่ง ตารางบนสุด คือ คำทั้งหมดในข้อความข้อเท็จจริง ส่วนตารางที่เหลือ คือ คำทั้งหมดในข้อความข้อกฎหมาย 66

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

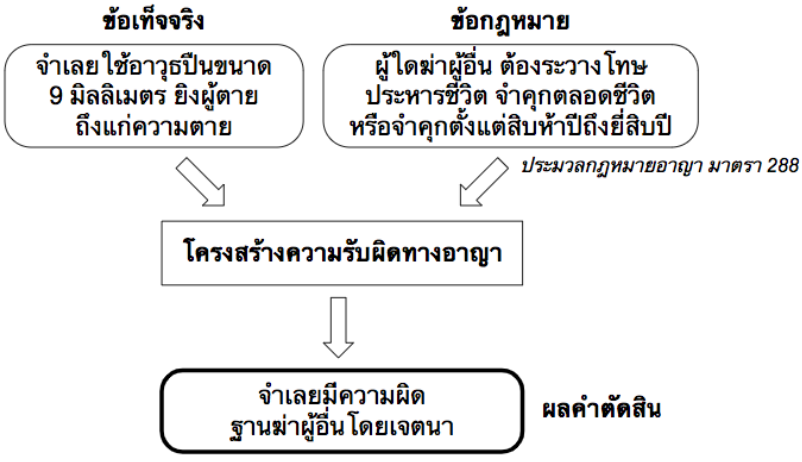
บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

กฎหมายอาญา คือ กฎหมายที่บัญญัติห้ามการกระทำบางอย่าง หรือบัญญัติให้มีการกระทำบางอย่าง โดยกำหนดบทลงโทษหากฝ่าฝืนเอาไว้ เช่น จำคุก ปรับ ฯลฯ กฎหมายอาญา ไม่ได้มีเพียงแค่ประมวลกฎหมายอาญา [1] เท่านั้น แต่ยังปรากฏอยู่ในกฎหมายฉบับอื่นซึ่งกำหนดบทลงโทษหากไม่ปฏิบัติตามกฎหมายฉบับนั้นไว้ด้วย ไม่ว่าจะเป็น กฎหมายจราจร กฎหมายอาวุธปืน เป็นต้น [2] ทั้งนี้ ประชาชนควรมีความรู้ความเข้าใจเบื้องต้นในกฎหมายอาญา เพื่อให้สามารถปฏิบัติตนได้อย่างถูกต้อง หลีกเลี่ยงผลกระทบต่อเสรีภาพ ชื่อเสียง หรือทรัพย์สิน ที่อาจเกิดขึ้นจากการถูกลงโทษทางอาญา นอกจากนี้ ยังเป็นไปตามสมมติฐานสำคัญข้อหนึ่งของกฎหมายอาญา นั่นคือ “ถือว่าทุกคนรู้กฎหมาย” โดยบุคคลจะอ้างว่าไม่รู้กฎหมายเพื่อให้ตนไม่ต้องรับผิดทางอาญาไม่ได้ ตามประมวลกฎหมายอาญา มาตรา 64 อีกด้วย

ในคดีอาญา ศาลจะใช้กฎหมายอาญาเพื่อวินิจฉัยว่า การกระทำของจำเลยเป็นความผิดตามกฎหมายหรือไม่ ถ้าเป็นความผิดแล้ว จำเลยจะต้องรับโทษหรือไม่เพียงใด โดยวิธีการใช้กฎหมายอาญานั้นจะเป็นไปตามโครงสร้างความรับผิดทางอาญา [2] ที่สกัดได้จากภาค 1 แห่งประมวลกฎหมายอาญา ทั้งนี้ ศาลจะนำข้อเท็จจริงที่รับฟังจากโจทก์และจำเลยมาปรับเข้ากับโครงสร้างข้างต้น ถ้าพบว่าการกระทำของจำเลยครบองค์ประกอบความผิดที่กฎหมายกำหนด ตลอดจน ไม่ปรากฏ เหตุยกเว้นความผิด และเหตุยกเว้นโทษตามกฎหมายแล้ว ศาลจึงจะพิพากษาลงโทษจำเลยได้ กระบวนการนี้พิจารณาได้ตามรูปที่ 1.1 ดังนั้น การใช้กฎหมายอาญาซับซ้อนในระดับหนึ่ง ไม่ใช่ เรื่องง่ายที่จะเข้าใจได้สำหรับผู้ ไม่มี ความรู้ทางกฎหมาย หรือแม้กระทั่งนักศึกษากฎหมายที่เริ่มต้นศึกษากฎหมายอาญา อนึ่ง ความรู้เกี่ยวกับโครงสร้างความรับผิดทางอาญา ได้แสดงรายละเอียดเอาไว้แล้วในหัวข้อที่ 2.3

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



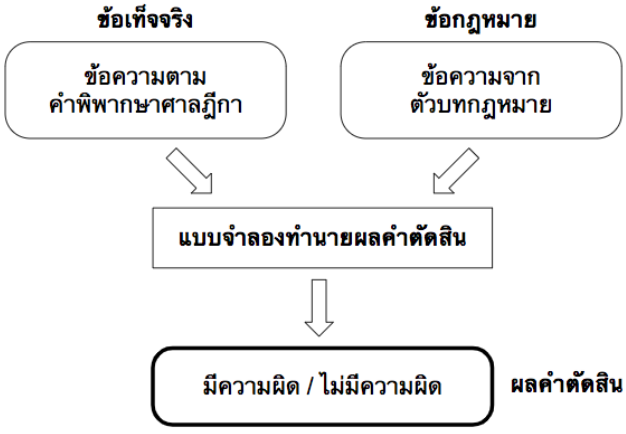
รูปที่ 1.1 *กระบวนการตัดสินคดีอาญาของผู้พิพากษาโดยใช้โครงสร้างความรับผิดทางอาญา*

ยิ่งไปกว่านั้น การศึกษากฎหมายอาญาของไทย ไม่ได้ ใช้เครื่องมือเพียงแค่ประมวลกฎหมายอาญาเท่านั้น แต่ยังมีคำพิพากษาของศาลฎีกาเป็นเครื่องมือสำคัญอีกชิ้นหนึ่งด้วย ทั้งนี้ แม้ว่าประเทศไทยจะใช้ระบบ กฎหมายแบบซิวิลลอว์ (Civil law) ซึ่งถือว่าคำพิพากษาของศาลเป็นเพียงตัวอย่างการใช้งานกฎหมายเท่านั้น มิใช่ตัวบทกฎหมายก็ตาม [3, 4] ทว่าคำพิพากษาศาลฎีกาก็เป็นคำตัดสินคดีขั้นเด็ดขาดจากศาลชั้นสูงสุด ถือเป็นตัวอย่างการใช้กฎหมายที่ชัดเจนและแน่นอน อีกทั้งหลักกฎหมายอาญาบางเรื่องก็เกิดขึ้นจากคำพิพากษาศาลฎีกาเอง ทั้งที่ไม่มีระบุไว้ในประมวลกฎหมายอาญาหรือกฎหมายอาญาฉบับอื่น ๆ มาก่อน เช่น หลักกฎหมายที่ว่า “การใช้อาวุธปืนถือว่ามีเจตนาฆ่า” [2] ดังนั้น การศึกษากฎหมายอาญาด้วยตัวบทกฎหมายและคำพิพากษาจึงซับซ้อนมากขึ้น

ด้วยเหตุนี้ ผู้วิจัยจึงสนใจพัฒนาแบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญาที่เรียนรู้จากคำพิพากษาศาลฎีกาไทยขึ้น โดยสามารถนำแบบจำลองนี้ไปประยุกต์ใช้ในงานดังต่อไปนี้ได้

1) เป็นเครื่องมือในการเรียนการสอนกฎหมายอาญา ทำให้นักศึกษากฎหมายเข้าใจโครงสร้างความรับผิดทางอาญา และหลักกฎหมายจากคำพิพากษาได้ง่ายยิ่งขึ้น สามารถใช้กฎหมายอาญาได้อย่างถูกต้อง
2) ใช้ ร่วมกับ ระบบสนับสนุนการตัดสินใจสำหรับนักกฎหมายเพื่อประโยชน์ในการดำเนินคดีอาญา โดยตรวจสอบได้ว่าจากข้อเท็จจริงที่รวบรวมมาและข้อกฎหมายที่เกี่ยวข้อง ผลคำตัดสินจะเป็นอย่างไร

ทั้งนี้ แบบจำลองที่ผู้วิจัยพัฒนาขึ้นจะทำนายผลคำตัดสินเฉพาะคดีอาญา โดยรับข้อมูลขาเข้าเป็นข้อเท็จจริงตามคำพิพากษาศาลฎีกาและข้อกฎหมายที่ใช้ในการตัดสิน จากนั้นให้ผลลัพธ์เป็นคำตัดสินว่าจำเลยมีความผิดตามข้อกฎหมายที่ระบุมาหรือไม่จากข้อเท็จจริงที่ให้มา โดยนำโครงสร้างความรับผิดทางอาญามาพิจารณาประกอบด้วย นอกจากนี้ แบบจำลองนี้ยังสามารถตอบคำถามในประเด็นต่าง ๆ ที่เกี่ยวข้องกับโครงสร้างความรับผิดทางอาญาได้ เช่น จำเลยมีการกระทำแล้วหรือไม่ กระทำการโดยเจตนาหรือไม่ ป้องกันตนโดยชอบด้วยกฎหมายหรือไม่ ฯลฯ ซึ่งลักษณะของแบบจำลองนี้จะตรงกับการจำแนกประเภทแบบสองคำตอบ (Binary classification) ที่ให้คำตอบอย่างใดอย่างหนึ่ง ระหว่าง ใช่ หรือ ไม่ใช่ นั่นเอง โครงสร้างของแบบจำลองนี้ พิจารณาได้ตามรูปที่ 1.2



รูปที่ 1.2 *โครงสร้างของแบบจำลองทำนายผลคำตัดสินคดีอาญาที่ผู้วิจัยพัฒนาขึ้น*

อย่างไรก็ดี จากผลงานวิจัยในอดีตที่เกี่ยวข้องกับการสร้างแบบจำลองทำนายผลคำพิพากษา พบว่า ข้อความที่นำมาใช้เป็นข้อมูลขาเข้าจะได้รับการแปลงเป็นตัวแทนข้อความ (Text representation) ในรูปเวกเตอร์ เพื่อนำไปใช้กับเทคนิคการเรียนรู้ต่าง ๆ ซึ่งการแปลงดังกล่าวนิยมใช้วิธีการดั้งเดิมของการประมวลภาษาธรรมชาติ เช่นแบบจำลองถุงคำ (Bag-of-Words model : BoW) แบบจำลองทีเอฟไอดีเอฟ (TF-IDF model ย่อมาจาก Term Frequency – Inverse Document Frequency model) หรือแบบจำลองเอ็นแกรม (N-Gram model) ฯลฯ ทว่าการสร้างเวกเตอร์ในลักษณะนี้ทำให้เกิดข้อจำกัดขึ้น กล่าวคือ

1) แบบจำลองเหล่านี้ ไม่สนใจ ลำดับของคำในข้อความทั้งข้อความ ทำให้สูญเสียบริบทของข้อความ ส่งผลให้ตัวแทนข้อความที่ได้นั้นยากต่อการวิเคราะห์ความหมายที่แท้จริง
2) ถ้าเอกสารทั้งหมดที่ใช้เป็นชุดข้อมูลสอนมีจำนวนคำในคลังศัพท์ (Vocabulary) มาก เวกเตอร์ของตัวแทนข้อความจะมีขนาดใหญ่มาก ทำให้ใช้พื้นที่หน่วยความจำมาก เช่น ถ้าข้อมูลคำพิพากษาทั้งหมดมีจำนวนคำในคลังศัพท์ทั้งหมด 100,000 คำ ก็อาจได้เวกเตอร์ที่มีขนาดถึง 100,000 มิติ

ด้วยเหตุนี้ ผู้วิจัยจึงสนใจพัฒนาแบบจำลองโดยใช้เทคนิคการเรียนรู้เชิงลึก (Deep learning) ด้วยการสร้างโครงข่ายประสาทเทียมที่ทำหน้าที่ 2 ส่วนในโครงข่ายเดียวกัน คือ การสร้างตัวแทนข้อความ และการทำนายผลคำตัดสิน ซึ่งเทคนิคนี้จะช่วยให้ตัวแทนข้อความในแบบจำลองมีคุณลักษณะสะท้อนความเป็นคำพิพากษา และข้อกฎหมายที่เกี่ยวข้องได้ เพราะโครงข่ายประสาทเทียมข้างต้นจะเรียนรู้การสร้างตัวแทนข้อความด้วยตัวเองโดยอาศัยการประมวลผลคำในข้อความตามลำดับทีละคำ อันเป็นการให้ความสำคัญกับลำดับของคำในข้อความนั้น ขณะเดียวกัน เทคนิคการเรียนรู้เชิงลึก ยังช่วยให้แบบจำลองสามารถเรียนรู้ข้อมูลสอนจากคำพิพากษาศาลฎีกาจำนวนมากได้ ไม่ว่าข้อมูลสอนทั้งหมดจะมีจำนวนคำมากเท่าใดก็ตาม เพราะสามารถกำหนดและควบคุมขนาดของโครงข่ายประสาทเทียมแต่ละชั้นได้

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

1.2 วัตถุประสงค์ของการวิจัย

เพื่อพัฒนาแบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญาที่เรียนรู้จากคำพิพากษาศาลฎีกาไทย โดยใช้เทคนิคการเรียนรู้เชิงลึก

1.3 ขอบเขตของการวิจัย

1) ข้อมูลในการวิจัยมาจากคำพิพากษาศาลฎีกาของประเทศไทยที่รวบรวมได้จากเว็บไซต์ระบบสืบค้นคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา [5] เฉพาะคำพิพากษาในคดีอาญา ทั้งนี้ คดีอาญาที่สนใจในงานวิจัยชิ้นนี้ คือ คดีที่ตัดสินตามประมวลกฎหมายอาญา ในฐานความผิดเกี่ยวกับชีวิตและร่างกาย ความผิดเกี่ยวกับชื่อเสียง และความผิดเกี่ยวกับทรัพย์

2) ข้อมูลขาเข้าของแบบจำลอง ประกอบด้วย
   2.1) ข้อความแสดงข้อเท็จจริงในคดี มีที่มาจากคำพิพากษาศาลฎีกา และ
   2.2) ข้อความแสดงข้อกฎหมายที่นำมาปรับใช้กับข้อเท็จจริง มีที่มาจากตัวบทกฎหมาย
3) ผลลัพธ์ของแบบจำลองมีลักษณะเป็นการจำแนกประเภทแบบสองคำตอบ โดยมี 2 รูปแบบ คือ
   3.1) ผลคำตัดสินว่าจำเลยมีความผิดตามข้อกฎหมายที่ระบุมาหรือไม่
   3.2) คำตอบในประเด็นต่าง ๆ ตามโครงสร้างความรับผิดทางอาญา เช่น จำเลยมีการกระทำหรือไม่ จำเลยกระทำความผิดโดยเจตนาหรือไม่ จำเลยป้องกันตนโดยชอบด้วยกฎหมายหรือไม่ จำเลยลงมือกระทำความผิดสำเร็จหรือไม่ ฯลฯ สำหรับประเด็นทั้งหมดที่แบบจำลองนี้ทำนายได้ระบุอยู่ในหัวข้อที่ 4.3.4
4) สำหรับการทำนายผลคำตัดสินว่าจำเลยมีความผิดตามกฎหมายหรือไม่นั้น จะไม่พิจารณาว่า จำเลยนั้นถูกตัดสินลงโทษสถานไหนเพียงใด กล่าวคือ ไม่สนใจว่า จำเลยถูกตัดสินจำคุกกี่ปี หรือถูกปรับเงินเท่าใด

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) สามารถประยุกต์ใช้เป็นเครื่องมือในการเรียนการสอนกฎหมายอาญา ทำให้นักศึกษากฎหมายเข้าใจโครงสร้างความรับผิดทางอาญา และหลักกฎหมายจากคำพิพากษาได้ง่ายยิ่งขึ้น สามารถใช้กฎหมายอาญาได้อย่างถูกต้อง
2) สามารถประยุกต์ใช้ร่วมกับระบบสนับสนุนการตัดสินใจสำหรับนักกฎหมายเพื่อประโยชน์ในการดำเนินคดีอาญา โดยตรวจสอบได้ว่าจากข้อเท็จจริงที่รวบรวมมา และข้อกฎหมายที่เกี่ยวข้อง ผลคำตัดสินเป็นอย่างไร

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

1.5 วิธีดำเนินการวิจัย

วิธีดำเนินการวิจัยแสดงเป็นขั้นตอนดังรายการต่อไปนี้ โดยมีแผนปฏิบัติงานตามวิธีดำเนินการวิจัยข้างต้นเป็นไปตามตารางที่ 1.1

1) ออกแบบและสร้างชุดข้อมูล
2) ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง
3) ออกแบบแนวคิด และวิธีการทดลอง
4) ทำการทดลอง
5) ศึกษางานวิจัยเพิ่มเติม
6) ปรับปรุงชุดข้อมูล แนวคิด และวิธีการทดลอง
7) สอบหัวข้อวิทยานิพนธ์
8) ปรับปรุงชุดข้อมูล และทำการทดลองเพิ่มเติม

9) สรุปผลการทดลอง
10) จัดทำบทความทางวิชาการ
11) จัดทำวิทยานิพนธ์
12) สอบวิทยานิพนธ์

1.6 ผลงานวิจัยที่ตีพิมพ์

" Predicting Judicial Decisions of Criminal Cases from Thai Supreme Court Using Bi-directional GRU with Attention Mechanism" โดย กานกวิญจน์ โค้วสีหวัฒน์ พีรพล เวทีกูล และปรัชญา บุญขวัญ ในงานประชุมวิชาการ "The 5th Asian Conference on Defense Technology (ACDT 2018)" จัดขึ้น ณ หอสมุด Ta Quang Buu ภายใน Hanoi University of Science and Technology (HUST) กรุงฮานอย ประเทศเวียดนาม ระหว่างวันที่ 25 - 26 ตุลาคม พ.ศ.2561

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 1.1 *ตารางแสดงแผนปฏิบัติงานตามวิธีดำเนินการวิจัย*

กิจกรรม	พ.ศ.2560			พ.ศ.2561											
	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.
1. ออกแบบและสร้างชุดข้อมูล															
2. ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง															
3. ออกแบบแนวคิด และวิธีการทดลอง															
4. ทำการทดลอง															
5. ศึกษางานวิจัยเพิ่มเติม															
6. ปรับปรุงชุดข้อมูล แนวคิดและวิธีการทดลอง															
7. สอบหัวข้อวิทยานิพนธ์															
8. ปรับปรุงชุดข้อมูล และทำการทดลองเพิ่มเติม															
9. สรุปผลการทดลอง															
10. จัดทำบทความทางวิชาการ															
11. จัดทำวิทยานิพนธ์															
12. สอบวิทยานิพนธ์															

บทที่ 2
แนวคิดและทฤษฎีเกี่ยวข้อง

2.1 ศาลฎีกาและคำพิพากษาศาลฎีกา

ศาลฎีกาเป็นศาลสูงสุดของประเทศไทย มีอำนาจพิจารณาพิพากษาคดีตามพระธรรมนูญศาลยุติธรรม [6] มาตรา 23 วรรคหนึ่ง ได้แก่ คดีที่รัฐธรรมนูญหรือกฎหมายบัญญัติให้เสนอต่อศาลฎีกาได้โดยตรง และคดีที่อุทธรณ์หรือฎีกาคำพิพากษาหรือคำสั่งของศาลชั้นต้น ศาลอุทธรณ์ หรือศาลอุทธรณ์ภาคตามที่กฎหมายบัญญัติ เว้นแต่กรณีที่ศาลฎีกาเห็นว่าข้อกฎหมายหรือข้อเท็จจริงที่อุทธรณ์หรือฎีกานั้นจะไม่เป็นสาระอันควรแก่การพิจารณา ศาลฎีกามีอำนาจไม่รับคดีไว้พิจารณาพิพากษาได้ ทั้งนี้ ตามระเบียบที่ที่ประชุมใหญ่ศาลฎีกากำหนดโดยประกาศในราชกิจจานุเบกษา อนึ่ง คดีที่ศาลฎีกาได้พิจารณาพิพากษาหรือมีคำสั่งแล้ว คู่ความไม่มีสิทธิที่จะทูลเกล้าฯ ถวายฎีกาคัดค้านคดีนั้นต่อไป กล่าวคือ คำพิพากษาของศาลฎีกาเป็นที่สุด ไม่อาจโต้แย้งใด ๆ ได้อีก

คำพิพากษาศาลฎีกาในคดีอาญา มีองค์ประกอบตามประมวลกฎหมายวิธีพิจารณาความอาญา [7] มาตรา 186 ประกอบมาตรา 215 และมาตรา 225 ได้แก่

1) ชื่อศาลและวันเดือนปี
2) คดีระหว่างใครโจทก์ใครจำเลย
3) เรื่อง
4) ข้อหาและคำให้การ
5) ข้อเท็จจริงซึ่งพิจารณาได้ความ
6) เหตุผลในการตัดสินทั้งในปัญหาข้อเท็จจริงและข้อกฎหมาย
7) บทมาตราที่ยกขึ้นปรับ
8) คำชี้ขาดให้ยกฟ้องหรือลงโทษ
9) คำวินิจฉัยของศาลในเรื่องของกลางหรือในเรื่องฟ้องทางแพ่ง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ด้วยเหตุที่คำพิพากษาศาลฎีกามีความสำคัญต่อการศึกษาค้นคว้าทางกฎหมายอย่างมาก จึงเกิดการรวบรวมและตีพิมพ์เผยแพร่คำพิพากษาศาลฎีกาเป็นประจำทุก ๆ ปี ทั้งนี้ หนังสือรวมคำพิพากษาศาลฎีกาที่ได้รับการอ้างอิงส่วนมากมาจากหน่วยงานของรัฐ 2 หน่วย ได้แก่ เนติบัณฑิตยสภา ซึ่งจัดทำหนังสือ “คำพิพากษาฎีกา” และสำนักงานศาลยุติธรรม ซึ่งจัดทำหนังสือ “คำพิพากษาศาลฎีกา” นอกจากนี้ ทางศาลฎีกายังได้จัดทำระบบสืบค้นคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา [5] เพื่อให้สามารถค้นหาคำพิพากษาศาลฎีกาผ่านทางเว็บไซต์ได้อีกด้วย โดยเนื้อหาคำพิพากษาที่ได้รับการเผยแพร่ผ่านช่องทางข้างต้น มักเป็นคำพิพากษาฉบับย่อ แบ่งได้เป็น 2 รูปแบบ คือ [8]

1) **ย่อสั้น** แสดงข้อกฎหมายที่น่าสนใจและสกัดได้จากเหตุผลในคำวินิจฉัยของศาลฎีกา และ
2) **ย่อยาว** แสดงรายละเอียดของคดีนั้นตั้งแต่ศาลชั้นต้น ศาลอุทธรณ์ มาจนถึงศาลฎีกาโดยสังเขป

2.2 ความรู้เบื้องต้นเกี่ยวกับกฎหมายอาญา

กฎหมายอาญา มีผู้ให้นิยามไว้มากมายดังปรากฏใน [9] แต่อาจกล่าวโดยสรุปได้ว่า กฎหมายอาญา คือ กฎหมายที่บัญญัติห้ามการกระทำบางอย่าง หรือบัญญัติให้มีการกระทำบางอย่าง โดยกำหนดบทลงโทษหากฝ่าฝืนเอาไว้เช่น ประมวลกฎหมายอาญา [1] มาตรา 288 บัญญัติว่า “ผู้ใดฆ่าผู้อื่น ต้องระวางโทษประหารชีวิต จำคุกตลอดชีวิต หรือจำคุกตั้งแต่สิบห้าปีถึงยี่สิบปี” ทั้งนี้ กฎหมายอาญา มิได้จำกัดอยู่แค่ประมวลกฎหมายอาญาเท่านั้น แต่ยังรวมถึงกฎหมายอื่น ๆ ซึ่งกำหนดความผิดและโทษเอาไว้ด้วย เช่น ความผิดเกี่ยวกับการพนันตาม พ.ร.บ.การพนัน พ.ศ.2478, ความผิดเกี่ยวกับกฎจราจรตาม พ.ร.บ.จราจรทางบก พ.ศ.2522, ความผิดเกี่ยวกับยาเสพติดตาม พ.ร.บ.ยาเสพติดให้โทษ พ.ศ.2522, ความผิดเกี่ยวกับการทารุณกรรมสัตว์ตาม พ.ร.บ.ป้องกันการทารุณกรรมและการจัดสวัสดิภาพสัตว์ พ.ศ.2557 ฯลฯ

เอกลักษณ์ของกฎหมายอาญา [2] ซึ่งได้รับการยอมรับในระดับสากล ปรากฏอยู่ในประมวลกฎหมายอาญา มาตรา 2 สรุปสาระสำคัญได้ว่า

1) ผู้กระทำไม่ต้องรับผิดในทางอาญา หากการกระทำนั้น ไม่มี กฎหมายบัญญัติไว้ในขณะกระทำว่าเป็นความผิดและกำหนดโทษไว้ เช่น การขับรถโดยประมาทชนรถยนต์ของผู้อื่นได้รับความเสียหาย กรณีนี้ ผู้กระทำไม่มีความผิดใด ๆ ตามกฎหมายอาญา เพราะไม่มีกฎหมายใดบัญญัติให้ การทำให้เสียทรัพย์โดยประมาท เป็นความผิด แต่ถ้าการขับรถนั้นเกิดจากเจตนา ผู้กระทำจึงจะมีความผิดและต้องรับโทษฐานทำให้เสียทรัพย์ ตามประมวลกฎหมายอาญา มาตรา 358
2) กฎหมายอาญาจะย้อนหลังให้ผลร้ายมิได้ หรือ “กฎหมายไม่มีผลย้อนหลัง” ตัวอย่างเช่น ปัจจุบันไม่มี กฎหมายบัญญัติให้ “การลักลอบเป็นชู้” เป็นความผิด ตามกฎหมายอาญา ถ้าต่อมามีกฎหมายบัญญัติให้การลักลอบเป็นชู้กลายเป็นความผิดและต้องรับโทษแล้ว บุคคลที่ลักลอบเป็นชู้ก่อนวันที่กฎหมายใหม่มีผลบังคับใช้ จะไม่ถือว่า มีความผิด เพราะกฎหมายจะไม่มีผลย้อนหลังกลับไปนั่นเอง อย่างไรก็ดี ถ้า กฎหมายใหม่มีผลเป็นคุณแก่ผู้กระทำความผิด กฎหมายนั้นมีผลย้อนหลังได้ เช่น ถ้ามีกฎหมายใหม่ออกมายกเลิกความผิดฐานหมิ่นประมาทตามประมวลกฎหมายอาญาแล้ว บุคคลที่กระทำความผิดอยู่ก่อนวันที่กฎหมายนี้ใช้บังคับ ย่อมได้รับประโยชน์ตามกฎหมายใหม่ด้วย ต้องถือว่าผู้นั้น ไม่มีความผิด อีกต่อไป
3) ถ้อยคำในกฎหมายอาญาจะต้องบัญญัติให้ชัดเจนแน่นอน ปราศจากความคลุมเครือ ตัวอย่างกฎหมายที่คลุมเครือ เช่นบัญญัติว่า ผู้ใด “หาเลี้ยงชีพด้วยวิธีการอันขัดต่อความสงบเรียบร้อยหรือศีลธรรมอันดีของประชาชน” มีความผิดฐานเป็นภัยสังคมและต้องรับโทษ กรณีนี้กฎหมายคลุมเครือ

เพราะประชาชนไม่อาจทราบได้ว่าการหาเลี้ยงชีพด้วยวิธีใด ขัดต่อความสงบเรียบร้อยหรือศีลธรรมอันดีของประชาชน

4) กฎหมายอาญาต้องตีความโดยเคร่งครัด ยกตัวอย่างเช่น การวางเพลิงเผาทรัพย์ของผู้อื่นเป็นความผิดตามประมวลกฎหมายอาญา มาตรา 217 แต่ถ้าบุคคลใดเผาทรัพย์ “ที่ตัวเองเป็นเจ้าของร่วมด้วย” แล้ว บุคคลนั้น ไม่มี ความผิดตามมาตรานี้ เพราะทรัพย์ที่มีผู้เผาเป็นเจ้าของร่วมด้วย ถือไม่ได้ว่าเป็นทรัพย์ของ “ผู้อื่น” ดังนั้น จะตีความขยายว่าทรัพย์ของผู้อื่น รวมถึง ทรัพย์ที่ผู้เผาเป็นเจ้าของรวมอยู่ด้วย ไม่ได้

2.3 การวินิจฉัยคดีอาญาด้วยโครงสร้างความรับผิดทางอาญา

ในการพิจารณาคดีอาญา ศาลจะนำข้อเท็จจริงที่รับฟังมาจากทั้งโจทก์และจำเลยมาปรับเข้ากับข้อกฎหมาย เพื่อตัดสินว่าการกระทำของจำเลยเป็นความผิดตามกฎหมายหรือไม่ ถ้าเป็นความผิดแล้ว จำเลยจะต้องรับโทษหรือไม่เพียงใด โดยวิธีการใช้กฎหมายอาญานั้นจะเป็นไปตามโครงสร้างความรับผิดในทางอาญา [2] ซึ่งสำคัญอย่างยิ่งเพราะจำเลยจะมีความผิดและต้องรับโทษทางอาญาได้**ก็ต่อเมื่อการกระทำของจำเลยนั้นเข้าองค์ประกอบทุกข้อในโครงสร้างนี้** หากขาดองค์ประกอบข้อใดข้อหนึ่งแม้เพียงข้อเดียวแล้ว จำเลยย่อมไม่มีความผิดและไม่ต้องรับโทษทางอาญา ส่งผลให้ศาลต้องพิพากษายกฟ้องโจทก์ปล่อยจำเลยไป ทั้งนี้ โครงสร้างความรับผิดทางอาญามีองค์ประกอบดังต่อไปนี้

1) **การกระทำครบ “องค์ประกอบ” ที่กฎหมายบัญญัติ** มีองค์ประกอบย่อย 4 ประการ ได้แก่

1.1) มีการกระทำ (Act) กล่าวคือ การเคลื่อนไหวหรือไม่เคลื่อนไหวร่างกายโดยรู้สำนึก เช่น การชกต่อยผู้อื่น แสดงว่ามีการกระทำเพราะเคลื่อนไหวร่างกายโดยใช้มือชกและรู้ตัวด้วยว่าตนกำลังชกผู้อื่นอยู่ แต่ถ้าเป็นการชกเพราะละเมอ จะถือว่า ไม่มี การกระทำ เพราะเคลื่อนไหวร่างกายโดยที่ไม่รู้ตัวว่ากำลังชกผู้อื่นอยู่

1.2) การกระทำนั้นครบองค์ประกอบภายนอก (External elements) ของความผิดในเรื่องนั้น ๆ กล่าวคือ การกระทำเป็นไปตามข้อกฎหมายที่กำลังสนใจอยู่ ซึ่งข้อกฎหมายเหล่านั้นมักระบุองค์ประกอบภายนอกไว้เป็น 3 ส่วน คือ ผู้กระทำ การกระทำ และวัตถุแห่งการกระทำ อาทิ

- ความผิดฐานฆ่าผู้อื่น องค์ประกอบภายนอก คือ ผู้กระทำ “ฆ่า” บุคคลอื่น
- ความผิดฐานลักทรัพย์ องค์ประกอบภายนอก คือ ผู้กระทำ “เอาไป” ซึ่งทรัพย์ของผู้อื่นหรือที่ผู้อื่นเป็นเจ้าของรวมอยู่ด้วย
- ความผิดฐานดูหมิ่น องค์ประกอบภายนอก คือ ผู้กระทำ “ดูหมิ่น” ผู้อื่น ในลักษณะซึ่งหน้าหรือด้วยการโฆษณา

1.3) **การกระทำนั้นครบองค์ประกอบภายใน (Internal/Mental elements) ของความผิดในเรื่องนั้น ๆ** กล่าวคือ ผู้กระทำมีเจตนาหรือกระทำโดยประมาท อย่างไรก็ดี ความผิดบางฐานก็อาจกำหนดให้ผู้กระทำต้องรับผิดได้แม้ ไม่มี เจตนาและ ไม่ ประมาท เลย ความผิดลักษณะนี้เรียกว่า ความผิดโดยเด็ดขาด (Strict liability) เช่น ความผิดฐานทำให้เกิดปฏิกูลแก่น้ำในบ่อหรือสระตามประมวลกฎหมายอาญา มาตรา 380 อนึ่ง การกระทำโดยเจตนา โดยประมาท หรือความผิดโดยเด็ดขาดนั้นล้วนเป็นเหตุการณ์ที่ไม่เกิดร่วมกัน (Mutually exclusive events)

1.4) **ผลของการกระทำที่เกิดขึ้นสัมพันธ์กับการกระทำ** ตามหลักในเรื่อง**ความสัมพันธ์ระหว่างการกระทำและผล (Causation)** เช่น จำเลยยิงผู้ตาย ผู้ตายตายด้วยกระสุนปืนของจำเลย แสดงว่าความตายนั้นสัมพันธ์กับการยิงปืนของจำเลย ทว่าความสัมพันธ์ระหว่างการกระทำกับผลนี้ พิจารณาเฉพาะฐานความผิดที่ต้องมีผลปรากฏเท่านั้น เช่น ความผิดฐานฆ่าผู้อื่นหรือทำร้ายร่างกาย ถ้าเป็นความผิดที่ผลสำเร็จ ไม่ใช่ สาระสำคัญ ก็ไม่ต้องพิจารณาข้อนี้ อาทิ ความผิดฐานลักทรัพย์ ถ้าพาทรัพย์ของผู้อื่นเคลื่อนที่ไปโดยทุจริต ความผิดก็สำเร็จแล้ว ไม่ว่าจะพาไปสำเร็จหรือไม่ก็ตาม

2) **การกระทำนั้น ไม่มี กฎหมายยกเว้นความผิด (Justification)** ทั้งนี้ กฎหมายยกเว้นความผิด เช่น การป้องกันโดยชอบด้วยกฎหมายตามประมวลกฎหมายอาญา มาตรา 68 หรือการแสดงความคิดเห็นหรือข้อความโดยสุจริตซึ่งยกเว้นความผิดฐานหมิ่นประมาท ตามมาตรา 329

3) **การกระทำนั้น ไม่มี กฎหมายยกเว้นโทษ (Legal excuse)** ทั้งนี้ กฎหมายยกเว้นโทษ เช่น การกระทำความผิดด้วยความจำเป็นตามประมวลกฎหมายอาญา มาตรา 67 หรือการพิสูจน์ข้อความที่หาว่าหมิ่นประมาทว่าเป็นความจริง ซึ่งถ้าจริงแล้วกฎหมายยกเว้นโทษตามมาตรา 330 วรรคหนึ่ง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ทั้งนี้ โครงสร้างความรับผิดทางอาญาข้างต้นสามารถสกัดเป็นชุดของกฎทางตรรกะได้ ก่อนอื่นขอกำหนดตัวแปรแสดงประพจน์ที่เกี่ยวข้องตามตารางที่ 2.1

ตารางที่ 2.1 *ตารางแสดงประพจน์ที่พิจารณาได้จากโครงสร้างความรับผิดทางอาญา*

ประพจน์	คำอธิบายประพจน์
isguilty	บุคคลจะต้องรับผิดทางอาญาหรือไม่
iscompleteelements	การกระทำครบองค์ประกอบความผิดหรือไม่
isjustify	การกระทำมีเหตุยกเว้น ความผิด หรือไม่
isexcuse	การกระทำมีเหตุยกเว้น โทษ หรือไม่
isact	มีการกระทำหรือไม่

ประพจน์	คำอธิบายประพจน์
$isexternalelements$	การกระทำครบองค์ประกอบ ภายนอก ของความผิดหรือไม่
$isinternalelements$	การกระทำครบองค์ประกอบ ภายใน ของความผิดหรือไม่
$iscause$	การกระทำสัมพันธ์กับผลแห่งการกระทำหรือไม่
$isintent$	การกระทำนั้น มีเจตนา หรือไม่
$isneglect$	การกระทำนั้น ประมาท หรือไม่
$isstrict$	การกระทำนั้น ต้องรับผิดโดยเด็ดขาด หรือไม่

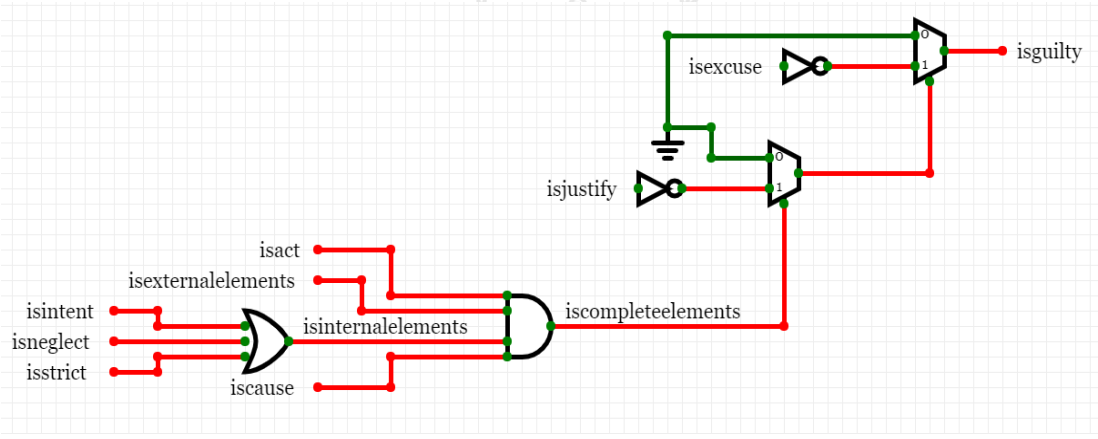
จากประพจน์ตามตารางที่ 2.1 สามารถแสดงกฎทางตรรกะได้ดังสมการต่อไปนี้

$$isguilty = iscompleteelements \wedge \neg isjustify \wedge \neg isexcuse \quad (2.1)$$

$$iscompleteelements = isact \wedge isexternalelements \wedge isinternalelements \wedge iscause \quad (2.2)$$

$$isinternalelements = isintent \vee isneglect \vee isstrict \quad (2.3)$$

นอกจากนี้ กฎทางตรรกะข้างต้นยังสามารถแสดงเป็นวงจรของลอจิกเกตได้ดังรูปที่ 2.1



รูปที่ 2.1 *ลอจิกเกตซึ่งเกิดจากกฎทางตรรกะแสดงโครงสร้างความรับผิดทางอาญา*

เนื่องด้วยการวินิจฉัยคดีอาญาส่วนใหญ่มักเป็นการตัดสินว่าจำเลยกระทำความผิดอาญาตามฟ้องจริงหรือไม่ ดังนั้น $isguilty$ จึงถือเป็น **ผลคำตัดสิน** ในคดีอาญาได้ ส่วนประพจน์อื่น ๆ ซึ่งแสดงองค์ประกอบต่าง ๆ ตามโครงสร้างความรับผิดทางอาญาถือเป็น **ประเด็น** ในคดีอาญา ประเด็นเหล่านี้จะนำไปสู่ผลคำตัดสินในท้ายที่สุด

อย่างไรก็ดี เนื่องด้วยสมการ (2.1) และ (2.2) เป็นกฎทางตรรกะที่เกิดจากประพจน์เชื่อม (Conjunction) ทั้งหมด ดังนั้น ถ้าประพจน์แสดงประเด็นใดในคดีอาญามีค่าความจริงเป็น เท็จ แล้ว ผลคำตัดสินย่อมออกมาว่าจำเลยไม่มีความผิดเลย *โดยไม่ต้องพิจารณาประเด็นอื่นอีก* ยกตัวอย่างเช่น ในคดีลักทรัพย์ ถ้าพบว่าจำเลย “ยังไม่ได้เอาไป” ซึ่งทรัพย์ของผู้อื่นหรือที่ผู้อื่นเป็นเจ้าของรวมอยู่ด้วยแล้ว กรณีนี้จึงเป็นการขาดองค์ประกอบภายนอกของความผิดฐานลักทรัพย์ ส่งผลให้ $isexternalelements$ เป็น เท็จ และ $isguilty$ เป็น เท็จ สรุปว่าจำเลยไม่มีความผิดฐานลักทรัพย์

ตัวอย่างการวินิจฉัยคดีโดยอาศัยโครงสร้างความรับผิดทางอาญา เช่น ศาลฟังข้อเท็จจริงได้ว่าจำเลยใช้อาวุธปืนยิงผู้ตายด้วยเจตนาฆ่า กระสุนตัดขั้วหัวใจผู้ตาย เป็นเหตุให้ถึงแก่ความตาย และ โจทก์ฟ้องขอให้ลงโทษจำเลยในความผิดฐานฆ่าผู้อื่นตามประมวลกฎหมายอาญา มาตรา 288 ซึ่งบัญญัติว่า “ผู้ใดฆ่าผู้อื่น ต้องระวางโทษ...” เมื่อพิจารณาประกอบกับโครงสร้างความรับผิดในทางอาญาแล้ว พบว่า

1) การกระทำครบ “องค์ประกอบ” ที่กฎหมายบัญญัติ เพราะ
   1.1) มีการกระทำ คือ จำเลยใช้อาวุธปืนยิงผู้ตาย
   1.2) การกระทำครบองค์ประกอบ ภายนอก ของความผิดในเรื่องนั้น กล่าวคือ การกระทำของจำเลย เป็นการ “ฆ่าผู้ตาย” แล้ว
   1.3) การกระทำครบองค์ประกอบ ภายใน ของความผิดในเรื่องนั้น นั่นคือ จำเลยมีเจตนาฆ่าผู้ตาย
   1.4) การกระทำสัมพันธ์กับผลที่เกิดขึ้น นั่นคือ กระสุนปืนจากจำเลยทำให้ผู้ตายถึงแก่ความตาย
2) การกระทำนั้น ไม่มี กฎหมายยกเว้นความผิด เพราะไม่ปรากฏว่าผู้ตายมีเหตุป้องกันตัว หรือเหตุอื่นใดที่จะยกเว้นความผิดฐานฆ่าผู้อื่นได้
3) การกระทำนั้น ไม่มี กฎหมายยกเว้นโทษ เพราะไม่ปรากฏว่าผู้ตายมีเหตุจำเป็น หรือเหตุอื่นใดที่จะยกเว้นโทษของตนได้

สรุปว่า จำเลยมีความผิดฐานฆ่าผู้อื่น ตามประมวลกฎหมายอาญา มาตรา 288

2.3.1 การพยายามกระทำความผิด

หลักการเรื่องการพยายามกระทำความผิดได้รับการบัญญัติอยู่ในประมวลกฎหมายอาญา มาตรา 80 วรรคหนึ่ง โดยประกอบด้วยเงื่อนไข 3 ประการคือ [2]

1) ผู้กระทำต้องมี **เจตนา** กระทำความผิด (ดังนั้น จึงไม่มีเรื่องพยายามกระทำโดยประมาท)

2) ผู้กระทำจะต้องกระทำการเพื่อให้บรรลุตามเจตนา อันเป็นการกระทำที่เลยขั้นตระเตรียม กล่าวคือ ถึงขั้น**ลงมือ**กระทำความผิดแล้ว

3) ผู้กระทำกระทำไป **ไม่ตลอด** หรือกระทำไปตลอดแล้ว แต่การกระทำนั้น **ไม่บรรลุผล**

ตัวอย่างการพยายามกระทำความผิด เช่น จำเลยต้องการฆ่าผู้เสียหาย จากนั้น

- จำเลยจ้องเล็งอาวุธปืนไปที่ผู้เสียหายแล้ว แต่มีผู้มาปัดปืนเสียก่อนทำให้ผู้เสียหายไม่ถูกยิง กรณีนี้จำเลยมีความผิดฐานพยายามฆ่าผู้เสียหาย เพราะมีเจตนาฆ่าและลงมือแล้วแต่กระทำการไป *ไม่ตลอด*
- จำเลยจ้องเล็งอาวุธปืนและยิงไปที่ผู้เสียหายแล้ว แต่ผู้เสียหายไม่ถึงแก่ความตาย เพียงแค่ได้รับบาดเจ็บเท่านั้น กรณีนี้จำเลยมีความผิดฐานพยายามฆ่าผู้เสียหายเช่นกัน เพราะมีเจตนาฆ่าและลงมือกระทำความผิดไปตลอดแล้วด้วยการยิงผู้เสียหาย เพียงแต่การฆ่า *ไม่บรรลุผล* เพราะผู้เสียหายไม่ถึงแก่ความตายสมดังเจตนาของจำเลย

นอกจากการพยายามกระทำความผิดตามมาตรา 80 แล้วยังมีการพยายามกระทำความผิดซึ่งเป็นไปไม่ได้โดยแน่แท้ตามมาตรา 81 วรรคหนึ่งด้วย โดยยังคงเงื่อนไขเดิมที่ได้กล่าวมาแล้ว แต่เพิ่มเงื่อนไขประการที่ 4 ขึ้นมา นั่นคือ “การกระทำนั้น ไม่สามารถบรรลุผลได้อย่างแน่แท้ เพราะเหตุ **ปัจจัย** ซึ่งใช้ในการกระทำ หรือเหตุแห่ง **วัตถุที่มุ่งหมายกระทำต่อ**” [2] ตัวอย่างการพยายามกระทำความผิดซึ่งเป็นไปไม่ได้โดยแน่แท้ เช่น จำเลยต้องการฆ่าผู้เสียหาย จากนั้น

- จำเลยใช้อาวุธปืนสั้นไทยประดิษฐ์ซึ่งตนประดิษฐ์ขึ้นเองยิงใส่ผู้เสียหาย แต่อาวุธปืนนี้มีอานุภาพต่ำไม่อาจฆ่าบุคคลใดให้ตายได้ ผู้เสียหายจึงได้รับบาดเจ็บเป็นเพียงรอยช้ำแดงบนผิวหนังเท่านั้น เช่นนี้จำเลยมีความผิดฐานพยายามฆ่าผู้อื่นซึ่งเป็นไปไม่ได้โดยแน่แท้ *เพราะเหตุปัจจัยซึ่งใช้ในการกระทำ* นั่นคืออาวุธที่ใช้ในการฆ่า ไม่มี อานุภาพร้ายแรงเพียงพอที่จะฆ่าผู้อื่นให้ถึงแก่ความตายได้
- จำเลยลอบเข้ามาที่บ้านผู้เสียหายแล้วใช้อาวุธปืนยิงใส่คนที่จำเลยคิดว่าเป็นผู้เสียหาย ทว่าแท้จริงแล้วกลับเป็นเพียงหุ่นที่ผู้เสียหายตั้งเอาไว้หลอกจำเลย เพราะรู้ตัวล่วงหน้าว่าจะถูกจำเลยลอบยิง เช่นนี้ จำเลยมีความผิดฐานพยายามฆ่าผู้อื่นซึ่งเป็นไปไม่ได้โดยแน่แท้ *เพราะเหตุแห่งวัตถุที่มุ่งหมายกระทำต่อ* นั่นคือจำเลยกระทำต่อหุ่น ไม่ใช่ตัวผู้เสียหาย จึงไม่มีทางที่จะฆ่าผู้เสียหายได้นั่นเอง

การพยายามกระทำความผิดทั้งสองรูปแบบ ทำให้ผู้กระทำความผิด **ได้รับโทษที่กฎหมายกำหนดไว้สำหรับความผิดนั้น ในรูปแบบอัตราส่วน** กล่าวคือ

- การพยายามกระทำความผิดตามมาตรา 80 ต้องระวางโทษสองในสามส่วน
- การพยายามกระทำความผิดซึ่งเป็นไปไม่ได้โดยแน่แท้ตามมาตรา 81 ต้องระวางโทษ ไม่เกิน กึ่งหนึ่ง เว้นแต่จะเป็นกรณีกระทำไปโดยความเชื่ออย่างงมงาย ซึ่งศาลจะไม่ลงโทษผู้นั้นก็ได้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อย่างไรก็ดี เนื่องด้วยการพยายามกระทำความผิด ไม่ได้ เป็นองค์ประกอบในโครงสร้างความรับผิดทางอาญาตาม [2] ดังนั้น เพื่อให้สอดคล้องกับการวิเคราะห์ความรับผิดทางอาญาตามโครงสร้างดังกล่าว การพยายามกระทำความผิด จึงอาจพิจารณาได้เป็น **ความล้มเหลวที่เกิดขึ้นจากการกระทำความผิด** โดยภายใต้เงื่อนไขที่ว่าบุคคลนั้นต้องรับผิดทางอาญา ($isguilty$ เป็นจริง) อยู่แล้ว มิเช่นนั้น ก็ไม่จำเป็นต้องวิเคราะห์การพยายามกระทำความผิดเลย ในเมื่อตัวจำเลยไม่มีความผิดหรือไม่ต้องรับโทษอยู่แล้ว

กำหนดให้ $isattempt$ เป็นประพจน์แสดงการพยายามกระทำความผิด

$isfail$ เป็นประพจน์แสดงความล้มเหลวที่เกิดขึ้นจากการกระทำความผิด

สามารถแสดงกฎทางตรรกะได้ตามสมการต่อไปนี้

$$isattempt \quad = \quad isfail \wedge isguilty \tag{2.4}$$

อนึ่ง การพยายามกระทำความผิดซึ่งเป็นไปไม่ได้โดยแน่แท้ สามารถได้คำนวณได้ในลักษณะเดียวกัน

กำหนดให้ $isimpossibleattempt$ เป็นประพจน์แสดงการพยายามกระทำความผิดซึ่งเป็นไปไม่ได้โดยแน่แท้

$isalwaysfail$ เป็นประพจน์แสดงความล้มเหลวอย่างแน่นอนในการกระทำความผิด

สามารถแสดงกฎทางตรรกะได้ตามสมการต่อไปนี้

$$isimpossibleattempt \quad = \quad isalwaysfail \wedge isguilty \tag{2.5}$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.4 ตัวแทนข้อความ (Text representation)

ในการสร้างระบบที่ใช้องค์ความรู้การประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) นั้น เราต้องทำให้ข้อความกลายเป็น **ข้อมูลตัวแทน (Representation / Feature)** เพื่อให้สามารถประมวลผลกับเทคนิคการเรียนรู้ต่าง ๆ ต่อไปได้ เราเรียกขั้นตอนนี้ว่าการสร้าง ตัวแทนข้อความ

ในการสร้างตัวแทนข้อความนั้น เรามักพิจารณาข้อความในหน่วยย่อย ซึ่งอาจเป็นระดับคำ หรือระดับตัวอักษรก็ได้ แต่ในที่นี้ จะเน้นพิจารณาที่ระดับคำเป็นหลัก โดยเราจะแปลงคำแต่ละคำให้กลายเป็นเวกเตอร์ แล้วนำเวกเตอร์ของคำทั้งหมดมาต่อกันจนกลายเป็นเมทริกซ์ของคำทั้งหมดที่ปรากฏในข้อความนั้นตามลำดับ และถือได้ว่าเป็นตัวแทนของข้อความนั้นแล้ว ทั้งนี้ กำหนดให้

t_1 แทนข้อความ “ผู้ตายต่อว่าจำเลยอย่างรุนแรง จำเลยจึงยิงผู้ตายถึงแก่ความตาย”

t_2 แทนข้อความ “จำเลยถูกผู้ตายยิงก่อน จำเลยจึงยิงสวนถูกผู้ตายถึงแก่ความตาย”

V แทนคลังศัพท์ (Vocabulary) หรือพจนานุกรม (Dictionary) ของคำจากทุกข้อความที่จะประมวลผล

ทั้งนี้ V ประกอบด้วยคำทั้งหมดจาก t_1 และ t_2 รวม 11 คำ ได้แก่

[ผู้ตาย, ต่อว่า, จำเลย, อย่าง, รุนแรง, จึง, ยิง, ถึงแก่ความตาย, ถูก, ก่อน, สวน]

สำหรับวิธีการแทนข้อความจะหยิบยกมาบางส่วนดังนี้ โดยจะยกตัวอย่างกับ t_1 และ t_2

1) แบบจำลองถุงคำ เป็นการแทนข้อความในรูปของเวกเตอร์ขนาดเท่ากับจำนวนคำทั้งหมดในคลังศัพท์ โดยไม่สนใจ ไวยากรณ์และลำดับของคำในข้อความนั้น ทั้งนี้ ค่าของแต่ละมิติในเวกเตอร์จะเป็นความถี่ของคำดังกล่าวที่ปรากฏในข้อความนั้น ดังนั้น จากข้อความที่กำหนดไว้

t_1 แทนข้อความได้เป็น $[2 \quad 1 \quad 2 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0]$

t_2 แทนข้อความได้เป็น $[2 \quad 0 \quad 2 \quad 0 \quad 0 \quad 1 \quad 2 \quad 1 \quad 2 \quad 1 \quad 1]$

2) แบบจำลองทีเอฟไอดีเอฟ เป็นการแทนข้อความที่มีพื้นฐานมาจากถุงคำ แต่ค่าของสมาชิกในเวกเตอร์ทั้งหมดจะเกิดจากการคูณค่าความถี่ของคำที่ปรากฏในข้อความ กับค่าผกผันของความถี่ของคำจากทุกข้อความ ค่าของแต่ละมิติในทีเอฟไอดีเอฟสามารถคำนวณได้จากสมการต่อไปนี้

$$tfidf = tf \times idf \tag{2.6}$$

$$idf = \log_{10}\left(\frac{N}{n_t}\right) \tag{2.7}$$

โดยที่ $tfidf$ เป็นค่าของสมาชิกแต่ละตัว

tf เป็นค่าความถี่ของคำที่สนใจ

idf เป็นค่าผกผันของความถี่ของคำจากทุกข้อความ

N คือ จำนวนข้อความทั้งหมด

n_t คือ จำนวนข้อความที่คำที่สนใจนั้นปรากฏอยู่

ดังนั้น จากข้อความที่กำหนดไว้

t_1 แทนข้อความได้เป็น $[0 \quad 0.3 \quad 0 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0 \quad 0 \quad 0]$

t_2 แทนข้อความได้เป็น $[0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0.3 \quad 0 \quad 0.3 \quad 0 \quad 0.3 \quad 0.3]$

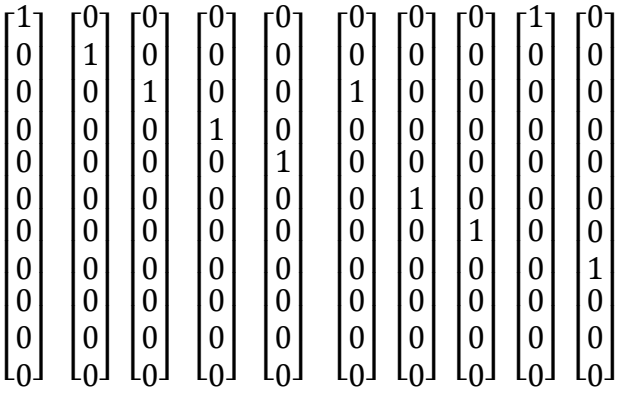
3) เวกเตอร์วัน - ฮอท (One-hot vector) คือ เวกเตอร์ที่มิติแทบทั้งหมดมีค่าเป็น 0 แต่มีมิติเดียวที่มีค่าเป็น 1 เปรียบได้ดังแผงวงจรไฟที่มีหลอดไฟหลาย ๆ หลอด แต่สว่างเพียงหลอดเดียว ทั้งนี้ ในการสร้างตัวแทนข้อความ ขนาดของเวกเตอร์วัน - ฮอทจะเท่ากับขนาดของคลังศัพท์ โดยมิติซึ่งตรงกับลำดับที่ของคำที่สนใจในคลังศัพท์จะมีค่าเป็น 1 นอกนั้นจะเป็น 0 ทั้งหมด เช่น ตามคลังศัพท์ คำว่า “จำเลย” เป็นคำลำดับที่ 3 จาก 11 คำ ดังนั้น เวกเตอร์วัน - ฮอทของคำนี้จึงมีค่าเป็น

$$[0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$$

ดังนั้น หากข้อความมีจำนวนคำทั้งหมด N คำแล้ว ตัวแทนข้อความที่ใช้เวกเตอร์วัน - ฮอท จะประกอบด้วยเวกเตอร์ N เวกเตอร์ โดยเวกเตอร์แต่ละตัวคือเวกเตอร์วัน - ฮอทที่ใช้แทนคำ ๆ นั้นนั่นเอง สำหรับข้อความที่กำหนดไว้ ขอแสดงเวกเตอร์วัน - ฮอทเป็นแนวตั้ง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

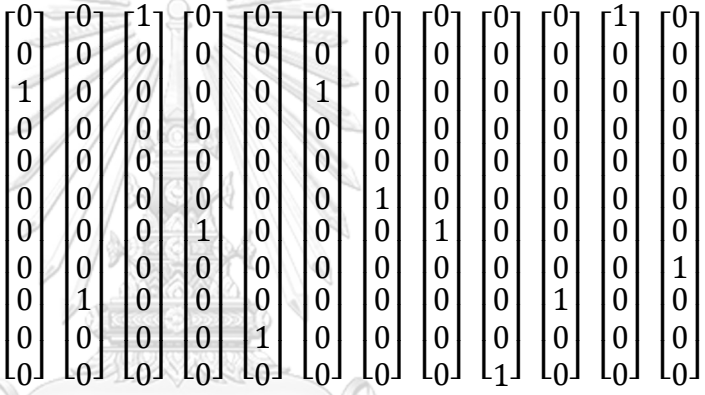
t_1 แทนข้อความได้เป็น



ผู้ตาย ต่อว่า จำเลย อย่าง รุนแรง จำเลย จึง ยิง ผู้ตาย ถึงแก่ความตาย

จะสังเกตได้ว่า คำว่า ผู้ตาย และ จำเลย มีความถี่เป็น 2 ดังนั้น จะมีเวกเตอร์วัน – ฮอทของคำดังกล่าวอยู่ 2 เวกเตอร์ ประจำตำแหน่งตามลำดับของคำใน t_1 ทั้งนี้ t_2 จะมีลักษณะอย่างเดียวกัน

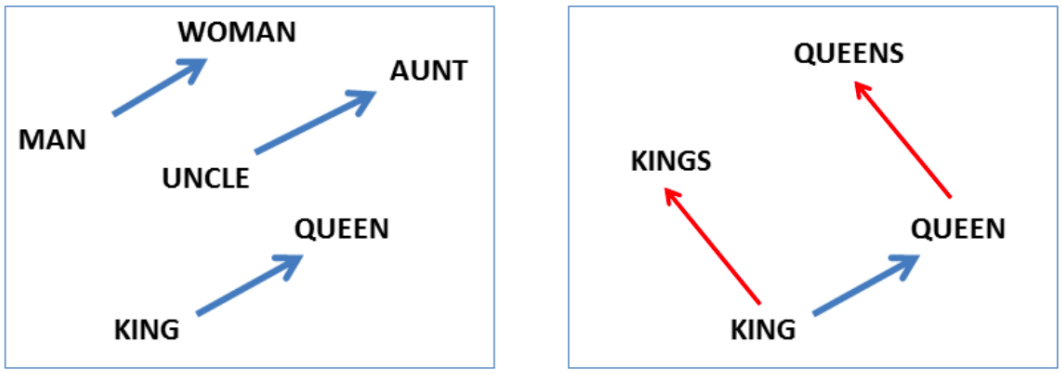
t_2 แทนข้อความได้เป็น



จำเลย ถูก ผู้ตาย ยิง ก่อน จำเลย จึง ยิง สวน ถูก ผู้ตาย ถึงแก่ความตาย

4) การฝังคำ (Word embedding) คือ การแทนคำด้วยเวกเตอร์หนึ่งขนาดเท่าใดก็ได้ เรียกว่า เวกเตอร์คำ (Word vector) โดยค่าแต่ละมิติมักเป็นค่าสุ่ม มีลักษณะคล้ายกับเวกเตอร์วัน - ฮอท ตรงที่คำแต่ละคำจะมีเวกเตอร์เป็นของตัวเองเพื่อใช้กับทุกเอกสาร และจำนวนเวกเตอร์ที่ใช้จะเท่ากับจำนวนคำทั้งหมดในข้อความ แต่การฝังคำมีข้อดีกว่า เพราะสามารถควบคุมขนาดเวกเตอร์คำตามต้องการได้ เช่น คลังศัพท์มีจำนวน 100,000 คำ แต่เวกเตอร์คำอาจมีขนาดมิติเพียงแค่ 100 เท่านั้นก็ได้ สำหรับเทคนิคการฝังคำนั้นมีมากมาย เช่น แอลเอสเอ (LSA ย่อมาจาก Latent Semantic Analysis) ตารางค้นหา (Lookup table) [10] เวิร์ดทูเวก (Word2Vec) [11] และโกลฟ (GloVe) [12]

ทั้งนี้ งานวิจัยด้านการฝังคำมีแนวโน้มทำให้เวกเตอร์คำสามารถสื่อความหมายทางภาษาได้ โดยสังเกตจากระยะห่างระหว่างเวกเตอร์ (Vector offset) [13] ดังตัวอย่างในรูปที่ 2.2



รูปที่ 2.2 *ตัวอย่างการสื่อความหมายทางภาษาของเวกเตอร์คำด้วยระยะห่างระหว่างเวกเตอร์* [13]

จากรูปที่ 2.2 จะเห็นได้ว่าในภาพซ้าย เวกเตอร์คำว่า ผู้ชาย (Man) กับ ผู้หญิง (Woman) ลุง (Uncle) กับ ป้า (Aunt) และ พระราชา (King) กับ พระราชินี (Queen) มีระยะห่างสื่อถึงความสัมพันธ์ระหว่างคำในเรื่องเพศชาย – หญิง ส่วนภาพขวา เวกเตอร์คำว่า พระราชา กับ พระราชาหลายพระองค์ (Kings) และ พระราชินี กับ พระราชินีหลายพระองค์ (Queens) มีระยะห่างสื่อถึงความสัมพันธ์ระหว่างคำในเรื่องความเป็นเอกพจน์ - พหูพจน์

การฝังคำจะนำชุดข้อมูลสอนซึ่งได้แก่ข้อความจากบทความหรือสื่อต่าง ๆ มาผ่านเทคนิคการฝังคำที่ได้กล่าวมาแล้ว โดยมุ่งให้คู่ของคำที่มีความหมายคล้ายกันมีระยะห่างระหว่างเวกเตอร์ใกล้เคียงกัน สมมติว่ามีเวกเตอร์คำขนาด 3 มิติที่ผ่านการฝังคำมาแล้ว สำหรับข้อความที่กำหนดไว้ สามารถแทนข้อความได้ดังนี้

t_1 แทนได้ว่า $\begin{bmatrix}0.51\\0.27\\0.82\end{bmatrix}\begin{bmatrix}0.05\\0.03\\0.90\end{bmatrix}\begin{bmatrix}0.22\\0.79\\0.11\end{bmatrix}\begin{bmatrix}0.73\\0.19\\0.26\end{bmatrix}\begin{bmatrix}0.21\\0.16\\0.06\end{bmatrix}\begin{bmatrix}0.22\\0.79\\0.11\end{bmatrix}\begin{bmatrix}0.90\\0.76\\0.51\end{bmatrix}\begin{bmatrix}0.48\\0.38\\0.96\end{bmatrix}\begin{bmatrix}0.51\\0.27\\0.82\end{bmatrix}\begin{bmatrix}0.96\\0.86\\0.82\end{bmatrix}$

ผู้ตาย ต่อว่า จำเลย อย่าง รุนแรง จำเลย จึง ยิง ผู้ตาย ถึงแก่ความตาย

t_2 แทนได้ว่า

$\begin{bmatrix}0.22\\0.79\\0.11\end{bmatrix}\begin{bmatrix}0.91\\0.15\\0.75\end{bmatrix}\begin{bmatrix}0.51\\0.27\\0.82\end{bmatrix}\begin{bmatrix}0.48\\0.38\\0.96\end{bmatrix}\begin{bmatrix}0.71\\0.15\\0.03\end{bmatrix}\begin{bmatrix}0.22\\0.79\\0.11\end{bmatrix}\begin{bmatrix}0.90\\0.76\\0.51\end{bmatrix}\begin{bmatrix}0.48\\0.38\\0.96\end{bmatrix}\begin{bmatrix}0.76\\0.10\\0.91\end{bmatrix}\begin{bmatrix}0.91\\0.15\\0.75\end{bmatrix}\begin{bmatrix}0.51\\0.27\\0.82\end{bmatrix}\begin{bmatrix}0.96\\0.86\\0.82\end{bmatrix}$

จำเลย ถูก ผู้ตาย ยิง ก่อน จำเลย จึง ยิง สวน ถูก ผู้ตาย ถึงแก่ความตาย

2.5 โครงข่ายประสาทเทียม (Artificial Neural Network : ANN)

โครงข่ายประสาทเทียม หรือในชื่ออื่นเช่น ข่ายงานประสาทเทียม และ นิวรัลเน็ตเวิร์ค (Neural network) เป็นเทคนิคการเรียนรู้หนึ่งในศาสตร์การเรียนรู้ของเครื่อง (Machine learning) ซึ่งมีที่มาจากการทำงานของเซลล์ประสาทในสมองของมนุษย์ ทั้งนี้ โครงข่ายประสาทเทียมใช้วิธีการเรียนรู้เชิงแนะนำ (Supervised learning) เนื่องจากชุดข้อมูลสอนโครงข่ายประสาทเทียมจะเป็นชุดข้อมูลที่ทราบคำตอบ (Class label) ล่วงหน้าแล้วนั่นเอง ในหัวข้อนี้ จะเริ่มนำเสนอจากหน่วยย่อยที่สุดของโครงข่ายประสาทเทียมก่อน แล้วจึงจะอธิบายตัวโครงข่ายรูปแบบต่าง ๆ ในภายหลัง

2.5.1 เพอร์เซปตรอน (Perceptron) [14]

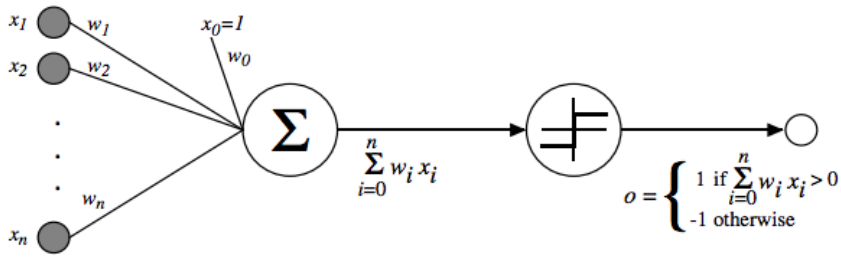
เพอร์เซปตรอน เป็นหน่วยย่อยที่สุดของโครงข่ายประสาทเทียม มีโครงสร้างที่พิจารณาได้ตามรูปที่ 2.3 ทั้งนี้ เพอร์เซปตรอนจะรับข้อมูลขาเข้าเป็นตัวเลขจำนวน n ค่า ตั้งแต่ $x_1, \dots, x_n$ จากนั้นจะไปผ่านขั้นตอนดังนี้

1) นำข้อมูลขาเข้าแต่ละค่าไปคูณถ่วงน้ำหนักด้วยค่าน้ำหนักของตัวเองซึ่งแทนด้วย $w_1, \dots, w_n$
2) นำค่าจากข้อ 1) แต่ละค่าไปบวกรวมกันพร้อมกับบวกด้วยค่าน้ำหนักพิเศษ w_0 หรือในตำราบางเล่มอาจเรียกว่าค่า b ซึ่งมาจากคำว่า ไบแอส (Bias) [15] ทั้งนี้ ค่าน้ำหนักพิเศษอาจมองเป็นข้อมูลขาเข้าตัวที่ 0 ซึ่งกำหนดตายตัวว่ามีค่าเป็น 1 เสมอก็ได้ กล่าวคือ $x_0 = 1$
3) นำค่าผลรวมจากข้อ 2) ไปเข้าฟังก์ชันกระตุ้น (Activation function) ซึ่งมีลักษณะเป็นฟังก์ชันขั้นบันได (Step function) ที่ให้ผลลัพธ์เป็นค่าใดค่าหนึ่งระหว่าง 1 กับ -1

ทั้งนี้ ผลลัพธ์จากข้อ 3) จะเป็นผลลัพธ์ของเพอร์เซปตรอนด้วยนั่นเอง ขั้นตอนทั้งหมดที่กล่าวถึงสามารถสรุปได้ในรูปแบบสมการดังต่อไปนี้

$$f(x) = \begin{cases} 1 & if\ w_0 + \sum_{i=1}^{n} w_i x_i > 0 \\ -1 & otherwise \end{cases} \tag{2.8}$$

อนึ่ง ฟังก์ชันขั้นบันไดที่ให้ผลลัพธ์เป็น 1 หรือ -1 นี้ นิยมเรียกว่า ฟังก์ชันสองขั้ว (Bipolar function) แต่หากฟังก์ชันขั้นบันไดนี้ให้ผลลัพธ์ระหว่าง 1 กับ 0 จะนิยมเรียกว่า ฟังก์ชันสองคำตอบ (Binary function)



รูปที่ 2.3 *โครงสร้างของเพอร์เซปตรอน*

การเรียนรู้ของเพอร์เซปตรอน จะมีกฎการสอนตามสมการดังต่อไปนี้

$$w_i \leftarrow w_i + \Delta w_i \tag{2.9}$$

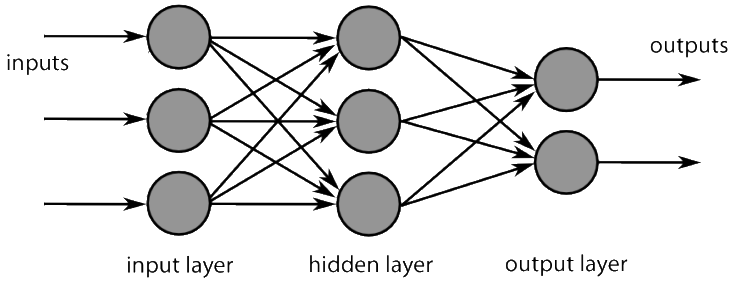
$$\Delta w_i = \eta(t - o)x_i \tag{2.10}$$

ทั้งนี้ w_i คือ น้ำหนักของเพอร์เซปตรอนตัวที่ i ซึ่งเราจะปรับ η ในสมการ (2.10) เรียกว่า อัตราการเรียนรู้ (Learning rate) เป็นค่าที่กำหนดว่าในการเรียนรู้แต่ละรอบจะเปลี่ยนแปลงน้ำหนักของข้อมูลขาเข้าทั้งหมดด้วยอัตราส่วนเท่าใด เมื่อเทียบกับผลต่างระหว่างคำตอบที่ถูกต้อง t กับผลลัพธ์ปัจจุบัน o

2.5.2 โครงข่ายประสาทเทียมแบบไปข้างหน้า (Feedforward neural network)

ในบางปัญหา เช่น การจำลองฟังก์ชันเอ็กซ์ออร์ (XOR หรือ Exclusive or) เพอร์เซปตรอนเพียงตัวเดียวไม่สามารถตอบคำถามได้ ดังนั้น จึงเป็นที่มาของโครงข่ายประสาทเทียมแบบไปข้างหน้า ซึ่งเป็นรูปแบบของโครงข่ายประสาทเทียมที่ง่ายที่สุด [15] ชื่อของโครงข่ายประสาทเทียมชนิดนี้ เกิดจากผลลัพธ์ของเพอร์เซปตรอนของแต่ละชั้นจะ ไม่ ย้อนกลับมาเป็นข้อมูลขาเข้าของชั้นก่อนหน้าอีก เป็นการส่งต่อข้อมูลแบบเดินหน้าทางเดียว สำหรับโครงสร้างของโครงข่ายประสาทเทียมชนิดนี้ พิจารณาได้จากรูปที่ 2.4 โดยแบ่งออกเป็น 3 ชั้น ประกอบด้วย

- ชั้นข้อมูลขาเข้า (Input layer) ยังคงเป็นค่าตัวเลขจำนวน n ค่า ตั้งแต่ $x_1, \ldots, x_n$ เหมือนเดิม
- ชั้นซ่อน (Hidden layer) ประกอบด้วย เพอร์เซปตรอนหลายตัว ซึ่งข้อมูลขาเข้าทุกตัวจากชั้นข้อมูลขาเข้าจะผ่านเข้าเพอร์เซปตรอนเหล่านี้ เพอร์เซปตรอนแต่ละตัวจะคำนวณผลลัพธ์ด้วยค่าน้ำหนักที่แตกต่างกันจนได้ผลลัพธ์ออกมา ผลลัพธ์จากเพอร์เซปตรอนทุกตัวในชั้นนี้ จะนำเข้าสู่ชั้นผลลัพธ์ต่อไป
- ชั้นผลลัพธ์ (Output layer) ประกอบด้วยเพอร์เซปตรอนอีกเช่นกัน และการทำงานคล้ายกับชั้นซ่อน แต่ข้อมูลขาเข้าของเพอร์เซปตรอนในชั้นนี้ คือ ผลลัพธ์จากเพอร์เซปตรอนแต่ละตัวในชั้นซ่อนนั่นเอง ทว่าเนื่องจากชั้นผลลัพธ์นี้จะให้ค่าเป็นผลลัพธ์ของโครงข่ายประสาทเทียมแบบไปข้างหน้า ดังนั้น จำนวนของเพอร์เซปตรอนในชั้นผลลัพธ์จะขึ้นอยู่กับลักษณะของปัญหาที่ใช้โครงข่ายรูปแบบนี้ในการแก้ปัญหาด้วย



รูปที่ 2.4 *โครงสร้างของโครงข่ายประสาทเทียมแบบไปข้างหน้า* [16]

สำหรับสมการที่เกิดขึ้นในโครงข่ายประสาทเทียมแบบไปข้างหน้าโดยใช้เพอร์เซปตรอนนั้น

กำหนดให้ h_i เป็น สมการของเพอร์เซปตรอนตัวที่ i ในชั้นซ่อน

p เป็นจำนวนเพอร์เซปตรอนในชั้นซ่อน และ

o_i เป็นสมการของเพอร์เซปตรอนตัวที่ i ในชั้นผลลัพธ์

สมการที่ได้คือ

$$h_i(x) = \begin{cases} 1 & if\ w_0 + \sum_{i=1}^{n} w_i x_i > 0 \\ -1 & otherwise \end{cases} \quad (2.11)$$

$$o_i(x) = \begin{cases} 1 & if\ w_0 + \sum_{i=1}^{p} w_i h_i > 0 \\ -1 & otherwise \end{cases} \quad (2.12)$$

แต่ในความเป็นจริงแล้ว แทบไม่มีใครใช้เพอร์เซปตรอนในการสร้างโครงข่ายประสาทเทียมแบบไปข้างหน้าเลย เพราะผลลัพธ์ที่ได้มีลักษณะเป็นเชิงเส้น และไม่สามารถตอบปัญหาที่ซับซ้อนขึ้นได้ การดัดแปลงเพอร์เซปตรอนจึงเกิดขึ้นโดยการเปลี่ยนฟังก์ชันกระตุ้นให้เป็นฟังก์ชันอื่น ฟังก์ชันกระตุ้นที่ได้รับความนิยม เช่น

1) ฟังก์ชันซิกมอยด์ (Sigmond function) สามารถแปลงค่าจำนวนจริงให้อยู่ในช่วงระหว่าง -1 ถึง 1 หรือ 0 ถึง 1 ได้ และมีคุณสมบัติเพิ่มคือ เมื่อนำไปหาอนุพันธ์แล้วจะไม่มีค่าเป็น 0 ดังนั้น จึงได้รับความนิยมในการเป็นฟังก์ชันประกอบการแพร่กระจายย้อนกลับซึ่งจะได้กล่าวต่อไป สำหรับสมการของฟังก์ชันนี้ คือ

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.13)$$

2) ฟังก์ชันไฮเพอร์โบลิกแทนเจนต์ (Hyperbolic tangent function หรือ tanh) สามารถแปลงค่าจำนวนจริงให้อยู่ในช่วงระหว่าง -1 ถึง 1 ได้ และมีคุณสมบัติหาอนุพันธ์แล้วไม่เป็น 0 เหมือนอย่างฟังก์ชันซิกมอยด์ สำหรับสมการของฟังก์ชันนี้ คือ

$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad (2.14)$$

3) ฟังก์ชันเรียงกระแสตรง (Rectifier) หรือที่นิยมเรียกอีกชื่อหนึ่งว่า ฟังก์ชันรีลู (ReLU ย่อมาจาก Rectified Linear Unit) จะแปลงค่าที่ติดลบให้กลายเป็น 0 ส่วนค่าที่ไม่ต่ำกว่า 0 จะมีค่าคงเดิม สมการมีดังนี้

$$ReLU(x) = \max(0, x) \quad (2.15)$$

4) ฟังก์ชันซอฟต์แม็กซ์ (Softmax function) จะแปลงค่าจำนวนจริงให้อยู่ระหว่าง 0 ถึง 1 มักนิยมใช้กับเวกเตอร์ โดยจะแปลงสมาชิกทุกตัวในเวกเตอร์ให้มีค่าระหว่าง 0 ถึง 1 ซึ่งเมื่อรวมผลลัพธ์จากฟังก์ชันซอฟต์แม็กซ์ทั้งหมดแล้ว จะได้ค่าเป็น 1 พอดี ด้วยเหตุนี้ ค่าผลลัพธ์แต่ละตัวจากฟังก์ชันนี้สามารถ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สะท้อนการกระจายของความน่าจะเป็นได้อีกด้วย สมมติว่ามีเวกเตอร์ขนาด k มิติ สมการของฟังก์ชันนี้สำหรับสมาชิกตัวที่ j คือ

$$softmax(x)_j = \frac{e^{x_j}}{\sum_{i=1}^{k} e^{x_i}} \quad (2.16)$$

หากฟังก์ชันกระตุ้นในเพอร์เซปตรอนจากฟังก์ชันขั้นบันไดเป็นฟังก์ชันใด ๆ ที่กล่าวมาข้างต้นแล้ว เราจะเรียกเพอร์เซปตรอนเหล่านั้นว่า หน่วยประสาทย่อย (Neural unit : Neuron) หรือ หน่วยย่อย (Unit) และสมการที่เกิดขึ้นในโครงข่ายประสาทเทียมแบบไปข้างหน้าโดยหน่วยย่อยเหล่านี้จะเปลี่ยนไป

กำหนดให้ h_i เป็นสมการของหน่วยย่อยตัวที่ i ในชั้นซ่อน

p เป็นจำนวนเพอร์เซปตรอนในชั้นซ่อน

o_i เป็นสมการของหน่วยย่อยตัวที่ i ในชั้นผลลัพธ์ และ

f เป็นฟังก์ชันกระตุ้นที่เลือกใช้

สมการที่ได้คือ

$$h_i(x) = f\left(w_0 + \sum_{i=1}^{n} w_i x_i\right) \quad (2.17)$$

$$o_i(x) = f\left(w_0 + \sum_{i=1}^{p} w_i h_i\right) \quad (2.18)$$

อนึ่ง สมการ (2.17) มักนำไปใช้ในงานโครงข่ายประสาทเทียมโดยทั่วไป และอาจเรียกได้อีกชื่อหนึ่งว่า ชั้นหนาแน่น (Dense layer) หรือชั้นเชื่อมโยงเต็มรูปแบบ (Fully connected layer)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.5.3 กฎการสอนโครงข่ายประสาทเทียมแบบไปข้างหน้า

ในการสอนโครงข่ายประสาทเทียมแบบไปข้างหน้าที่มีองค์ประกอบเป็นหน่วยย่อย (ไม่ใช่ เพอร์เซปตรอน) นั้น มักต้องมีค่าความผิดพลาด (Error) เกิดขึ้นเสมอ ดังนั้น กฎการสอนจะเป็นการลดค่าความผิดพลาดให้เหลือน้อยที่สุด กล่าวคือ เข้าใกล้ 0 ให้มากที่สุดนั่นเอง ทั้งนี้ เราจะมองค่าความผิดพลาดเป็นฟังก์ชันต้นทุน (Cost function) หรือฟังก์ชันการสูญเสีย (Loss function) เพื่อใช้สอนโครงข่าย

กำหนดให้ E คือค่าความผิดพลาด หรือฟังก์ชันการสูญเสีย

N คือจำนวนข้อมูลสอนทั้งหมด

t_i คือคำตอบที่ถูกต้องของข้อมูลสอนลำดับที่ i

o_i คือผลลัพธ์ปัจจุบันของข้อมูลสอนลำดับที่ i (ซึ่งจะต้องปรับให้ตรงกับคำตอบที่ถูกต้อง หากผลลัพธ์ยังผิดอยู่)

ค่าความผิดพลาดที่นิยมเลือกใช้มีดังนี้

1) ค่าเฉลี่ยความผิดพลาดยกกำลังสอง (Mean Squared Error : MSE)

$$\mathrm{E} = \frac{1}{2}\sum_{i=1}^{N}(t_i - o_i)^2 \tag{2.19}$$

2) ค่าเฉลี่ยครอส-เอนโทรปีแบบทวิภาค (Binary Cross-Entropy : BCE)

$$\mathrm{E} = -\frac{1}{N}\sum_{i=1}^{N} t_i \cdot \log(o_i) + (1 - t_i) \cdot \log(1 - o_i) \tag{2.20}$$

3) ค่าติดลบลอการิทึมของความเป็นไปได้ (Negative Log Likelihood : NLL) หากกำหนดให้ C เป็นจำนวนประเภทของคำตอบทั้งหมดที่เป็นไปได้ และ d_i คือผลต่างของความน่าจะเป็นระหว่างคำตอบที่ถูกต้องกับคำตอบทั้งหมดที่ทำนาย สามารถเขียนสมการฟังก์ชันการสูญเสียได้ว่า

$$\mathrm{E} = -\frac{1}{C}\sum_{i=1}^{C} \log d_i \tag{2.21}$$

ในการสอนโครงข่ายประสาทเทียมแบบไปข้างหน้า จะมีการหาค่าเหมาะที่สุด (Optimization) เพื่อมุ่งลดค่าความผิดพลาดให้ได้มากที่สุดในการสอนข้อมูลแต่ละรอบด้วย วิธีการที่นิยมใช้กันมากวิธีหนึ่ง คือ เอสจีดี (SGD ย่อมาจาก Stochastic Gradient Descent) โดยจะนำค่าจากฟังก์ชันการสูญเสียมาหาค่าเกรเดียนเทียบน้ำหนัก ทั้งนี้ ปรับน้ำหนักสำหรับข้อมูลขาเข้าลำดับที่ i ได้ดังนี้

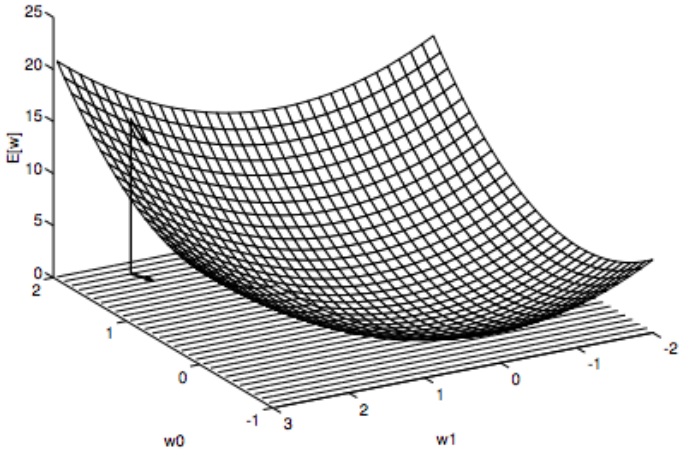
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$w_i \leftarrow w_i + \Delta w_i \tag{2.9}$$

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i} \tag{2.22}$$

ถ้าเรามีเพอร์เซปตรอนหรือหน่วยย่อยที่มีข้อมูลขาเข้า 1 ตัว และเรานำค่าน้ำหนักของข้อมูลขาเข้านั้นแทนด้วย w_1 ค่าน้ำหนักพิเศษ w_0 และค่าความผิดพลาด $\mathrm{E[w]}$ มาสร้างสเปซของความผิดพลาด จะพบว่าพื้นผิวของค่าความผิดพลาดมีลักษณะเป็นผิวโค้ง โดยค่าความผิดพลาดจะลดลงในลักษณะการลงเนินเขาเพื่อไปหาจุดต่ำที่สุดนั่นเอง พิจารณาได้จากกราฟในรูปที่ 2.5

อย่างไรก็ดี เนื่องจากค่าน้ำหนักที่ปรับเสร็จแล้วจะมีผล **เฉพาะ** ชั้นสุดท้ายของโครงข่ายเท่านั้น แต่เรามีชั้นของหน่วยย่อยถึง 2 ชั้น ดังนั้น เราจะต้องปรับค่าน้ำหนักทุกชั้นด้วยวิธีการแพร่กระจายย้อนกลับ (Backpropagation) สมมติว่าเราใช้ค่าความผิดพลาดเป็นค่าเฉลี่ยความผิดพลาดยกกำลังสอง [14]



รูปที่ 2.5 *สเปซแสดงค่าความผิดพลาดในการสอนเพอร์เซปตรอนหรือหน่วยย่อย เมื่อกำหนดค่าน้ำหนักของข้อมูลขาเข้า* w_1 *และค่าน้ำหนักพิเศษ* w_0 *แตกต่างกัน* [14]

กำหนดให้ x_{ij} คือ ข้อมูลขาเข้าลำดับที่ i ที่ใส่ในหน่วยย่อยลำดับที่ j

w_{ij} คือค่าน้ำหนักของข้อมูลขาเข้าลำดับที่ i ไปหาหน่วยย่อยลำดับที่ j

Δw_{ij} คือค่าน้ำหนักที่จะปรับเปลี่ยนของ w_{ij}

o_j คือผลลัพธ์ที่คำนวณได้จากหน่วยย่อยลำดับที่ j

t_j คือคำตอบที่ถูกต้องซึ่งหน่วยย่อยลำดับที่ j ต้องคำนวณออกมาให้ได้

การแพร่กระจายย้อนกลับในชั้นผลลัพธ์ k จะใช้สมการ (2.9) เป็นหลักโดยคำนวณเพิ่มเติมดังนี้

$$\Delta w_k = -\eta \frac{\partial E}{\partial w_k} = \eta \delta_k x_{jk} \quad (2.23)$$

$$\delta_k \leftarrow o_k(1-o_k)(t_k - o_k) \quad (2.24)$$

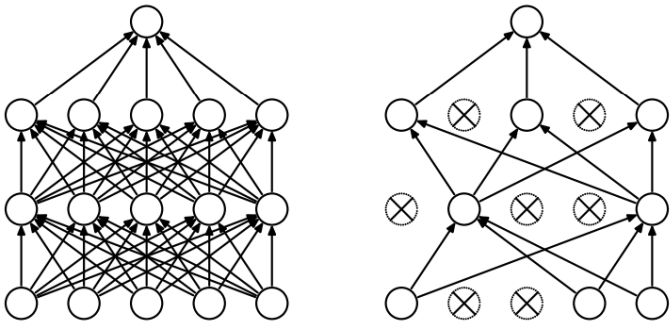
ส่วนการแพร่กระจายย้อนกลับในชั้นซ่อน j จะใช้สมการ (2.9) เป็นหลักโดยคำนวณเพิ่มเติมดังนี้

$$\Delta w_j = \eta \delta_j x_{ij} \quad (2.25)$$

$$\delta_j \leftarrow o_j(1-o_j) \sum_{k \in outputs} w_{jk}\delta_{jk} \quad (2.26)$$

2.5.4 การหยุดเรียนรู้กลางคัน (Drop-out) [17]

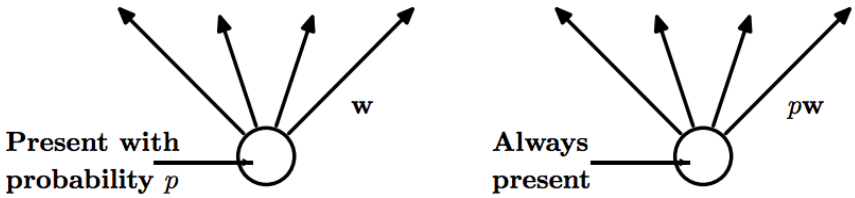
การหยุดเรียนรู้กลางคัน เป็นวิธีการป้องกัน ไม่ให้ โครงข่ายประสาทเทียมเรียนรู้โดยยึดติดกับข้อมูลสอนมากเกินไป (Overfitting) ด้วยวิธีสุ่มหยุดการทำงานของหน่วยย่อยในโครงข่ายระหว่างการเรียนรู้เป็นการชั่วคราว ทำให้หน่วยย่อยบางหน่วย ไม่มี การประมวลผลข้อมูล เปรียบเสมือนการหยุดหน่วยย่อยไม่ให้เรียนรู้กลางคัน ทั้งนี้ แนวคิดของวิธีการนี้แสดงตามรูปที่ 2.6 ซึ่งภาพซ้ายแสดงโครงข่ายประสาทเทียมที่มีชั้นซ่อน 2 ชั้น หน่วยย่อยแต่ละหน่วยจะถูกสุ่มหยุดเรียนรู้กลางคันด้วยค่าความน่าจะเป็นค่าหนึ่ง หากหน่วยย่อยใดถูกเลือกซึ่งแสดงเป็นวงกลมที่มีกากบาทภายในแล้ว ผลลัพธ์จากหน่วยย่อยก่อนหน้าจะ ไม่ เข้ามาประมวลผลในหน่วยย่อยนี้ ส่วนหน่วยย่อยอื่น ๆ จะประมวลผลไปตามปกติ



รูปที่ 2.6 *ภาพแสดงแนวคิดการหยุดเรียนรู้กลางคัน หน่วยย่อยที่ถูกหยุดเรียนรู้กลางคันแสดงตามภาพขวาด้วยวงกลมที่มีกากบาทภายใน* [17]

อย่างไรก็ตาม การหยุดเรียนรู้กลางคันจะทำในช่วงการสอนข้อมูลให้แบบจำลองเท่านั้น หากเป็นการทดสอบแบบจำลอง จะใช้วิธีถ่วงน้ำหนักเส้นทางของโครงข่ายแทนดังรูปที่ 2.7

CHULALONGKORN UNIVERSITY



รูปที่ 2.7 *ภาพซ้ายแสดงการหยุดการเรียนรู้กลางคันในช่วงสอนข้อมูลให้กับแบบจำลอง โดยสุ่มหยุดการทำงานของหน่วยย่อยด้วยค่าความน่าจะเป็น p ส่วนภาพขวาแสดงวิธีการใช้งานแบบจำลองในช่วงทดสอบข้อมูล โดย ไม่มี การหยุดการทำงานของหน่วยย่อยใด แต่จะถ่วงน้ำหนักเส้นทางของโครงข่ายที่จะออกไปหาหน่วยย่อยอื่นแทน* [17]

อนึ่ง ในกรณีโครงข่ายประสาทเทียมแบบวกกลับ (Recurrent Neural Network) การหยุดการเรียนรู้กลางคันจะใช้รูปแบบเดียวกันกับโครงข่ายประสาทเทียมทั่วไป แต่หน่วยย่อยที่ถูกหยุดจะเป็นหน่วยเดียวกันตลอดลำดับของข้อมูลเพื่อช่วยลดความผิดพลาดที่อาจเกิดจากการสุ่มหยุดหน่วยย่อยมากเกินไประหว่างช่วงประมวลผลลำดับข้อมูลนั้น [18]

2.6 การเรียนรู้เชิงลึก (Deep learning) [19]

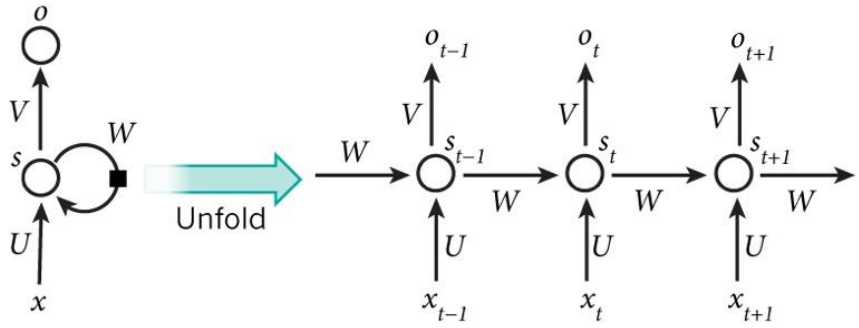
ความรู้ด้านโครงข่ายประสาทเทียมได้รับการวิจัยและพัฒนาจนเกิดเป็นเทคนิคการเรียนรู้เชิงลึกขึ้น โดยให้โครงข่ายประสาทเทียม **หลายชั้น** เรียนรู้ที่จะสร้างข้อมูลตัวแทนขึ้นมาเพื่อใช้ควบคู่กับอัลกอริทึมต่าง ๆ จากเดิมที่ต้องผ่านกระบวนการสกัดข้อมูลตัวแทน (Feature extraction) โดยมนุษย์ ซึ่งในเชิงการประมวลผลภาษาธรรมชาตินั้น เดิมเรามีวิธีสร้างตัวแทนข้อความรูปแบบต่าง ๆ เช่น ถุงคำ, ทีเอฟไอดีเอฟ, แอลเอสเอ ฯลฯ แต่สำหรับการเรียนรู้เชิงลึกแล้ว ตัวแทนข้อความจะเกิดจากการส่งข้อความเข้าไปในโครงข่ายประสาทเทียม **และให้โครงข่ายนั้นเรียนรู้เองว่าจะสร้างเวกเตอร์ตัวแทนข้อความอย่างไรให้สะท้อนข้อความนั้นได้มากที่สุด** ทั้งนี้ หากตัวแทนข้อความนั้นสามารถจำแนกออกเป็นหลายระดับได้ เราก็สามารถเพิ่มชั้นหน่วยย่อยของโครงข่ายประสาทเทียมเพื่อให้การเรียนรู้เชิงลึกทำได้ลึกขึ้น เช่น โครงข่ายประสาทเทียมที่เรียนรู้ข้อความผ่านชั้นหน่วยย่อยที่กำหนดตามโครงสร้างภาษา เริ่มตั้งแต่ระดับตัวอักษร พยางค์ คำ และวลี อย่างไรก็ดี ในผลงานวิจัยชิ้นนี้ จะสนใจเฉพาะการเรียนรู้ข้อความในระดับคำเท่านั้น

อนึ่ง เนื่องด้วยการเรียนรู้เชิงลึกอาศัยโครงข่ายประสาทเทียม ดังนั้น จึงอาจต่อชั้นหน่วยย่อยออกไปเพื่อเอาตัวแทนข้อความที่ได้ไปใช้ประมวลผลงานที่ต้องการได้ทันที เช่น การวิเคราะห์อารมณ์ความรู้สึก (Sentiment analysis), การแปลภาษา (Translation), การจำลองหัวข้อ (Topic modeling) หรือแม้แต่การจำแนกประเภทข้อความต่าง ๆ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.7 โครงข่ายประสาทเทียมแบบวกกลับ (Recurrent Neural Network : RNN) [19]

โครงข่ายประสาทเทียมแบบวกกลับ คือ โครงข่ายประสาทเทียมที่ประมวลผลข้อมูลขาเข้าตามลำดับ (Sequential) เช่น ลำดับของคำที่ประกอบรวมกันเป็นข้อความ หรือข้อมูลที่เข้ามาตามเวลา โครงสร้างของโครงข่ายประสาทเทียมแบบวกกลับนั้นคล้ายคลึงกับโครงข่ายประสาทเทียมทั่วไป แต่มีข้อมูลขาเข้าเพิ่มเติมนอกเหนือจากข้อมูลขาเข้า ณ เวลาปัจจุบัน คือ ผลลัพธ์จากชั้นซ่อนในช่วงเวลาก่อนหน้า พิจารณาโครงสร้างได้ตามรูปที่ 2.8



รูปที่ 2.8 *โครงสร้างของโครงข่ายประสาทเทียมแบบวกกลับ ซึ่งความจริงแล้วคือการนำโครงข่ายประสาทเทียมธรรมดามาต่อกันเพื่อรับข้อมูลที่มีลักษณะเป็นลำดับ หรือมาตามเวลานั่นเอง* [19]

กำหนดให้ s_t คือ ค่าสถานะของชั้นซ่อน (Hidden state) หรือผลลัพธ์ที่ช่วงเวลา t

W คือ ค่าของน้ำหนักที่ใช้คูณกับข้อมูลจากชั้นซ่อนเมื่อช่วงเวลา $t-1$

U คือ ค่าของน้ำหนักที่ใช้คูณกับข้อมูลนำเข้าที่ช่วงเวลา t

x_t คือ ค่าของข้อมูลนำเข้าที่ช่วงเวลา t

σ คือ ฟังก์ชันกระตุ้น

b คือ ค่าน้ำหนักพิเศษ หรือ ไบแอส

สมการของโครงข่ายประสาทเทียมแบบวกกลับโดยทั่วไป คือ

$$s_t = \sigma(Ws_{t-1} + Ux_t + b) \tag{2.27}$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ด้วยเหตุที่โครงข่ายประสาทเทียมแบบวกกลับรับข้อมูลขาเข้าเป็นผลลัพธ์จากชั้นซ่อนในช่วงเวลาก่อนหน้า s_{t-1} กับข้อมูลขาเข้า ณ เวลาปัจจุบัน x_t โครงข่ายประสาทเทียมแบบวกกลับจึงมีข้อได้เปรียบกว่าโครงข่ายประสาทเทียมเดิมในเชิงการเชื่อมโยงข้อมูลที่เรียนรู้ ณ เวลาปัจจุบันกับสิ่งที่เคยเรียนรู้มาแล้ว เปรียบได้ดังสมองของมนุษย์ที่สามารถรับรู้ข้อมูลที่ตนเรียนรู้มาในเวลาปัจจุบันแล้วเชื่อมโยงกับข้อมูลที่ตนจดจำและเคยเรียนรู้มาในอดีตได้ แตกต่างจากโครงข่ายประสาทเทียมเดิมที่สนใจข้อมูล ณ เวลาปัจจุบันเท่านั้น

2.7.1 ปัญหาการพึ่งพาระยะยาว (Long-term dependency)

การใช้โครงข่ายประสาทเทียมแบบวกกลับโดยตรง ไม่ได้รับความนิยมมากนักในการประมวลผลภาษาธรรมชาติ เนื่องจากปัญหาการพึ่งพาระยะยาว กล่าวคือ ถ้าข้อความที่นำมาใช้เป็นข้อมูลขาเข้ามีความยาวมาก เช่น บทความ หรือวรรณคดี – วรรณกรรมแล้ว โครงข่ายประสาทเทียมแบบวกกลับจะ ไม่สามารถ เก็บรักษาข้อมูลที่เคยเรียนรู้มาในอดีตทั้งหมดได้ เพราะเกิดปัญหาเกรเดียนเลือนหาย (Vanishing gradient problem) จากการ

แพร่กระจายย้อนกลับซึ่งค่าน้ำหนักของหน่วยย่อยบางหน่วยเปลี่ยนแปลงน้อยมาก จนแทบไม่มีการเปลี่ยนแปลงใดในหน่วยย่อยนั้น อาจกล่าวได้ว่าโครงข่ายประสาทเทียมแบบวกกลับเกิดอาการ “ลืม” ข้อมูลที่เคยเรียนรู้มาก่อนหน้าทำนองเดียวกับสมองของมนุษย์ที่อ่านวรรณคดีเรื่องยาวต่อเนื่องเป็นเวลานาน จนลืมเหตุการณ์ในตอนก่อนหน้า

ด้วยเหตุนี้ จึงเกิดการดัดแปลงโครงข่ายประสาทเทียมแบบวกกลับ ให้สามารถ “จดจำ” ข้อมูลที่เคยเรียนรู้มาในอดีตไว้ได้ โดยใช้แนวคิดของ “หน่วยความจำ” (Memory) เข้ามาประกอบ ขอแสดงตัวอย่างโครงข่ายประสาทเทียมแบบวกกลับที่เสริมหน่วยความจำ 2 โครงข่าย ดังต่อไปนี้

2.7.2 โครงข่ายความจำระยะสั้นขนาดยาว (Long Short-Term Memory : LSTM) [20, 21]

โครงข่ายความจำระยะสั้นขนาดยาวสร้างขึ้นตั้งแต่ค.ศ.1997 และได้รับการปรับปรุงเมื่อค.ศ.2000 โดยรับข้อมูลขาเข้าเป็นผลลัพธ์จากชั้นซ่อนในช่วงเวลาก่อนหน้า s_{t-1} กับข้อมูลขาเข้า ณ เวลาปัจจุบัน x_t ทั้งนี้เอกลักษณ์สำคัญ ของโครงข่ายชนิดนี้ คือ เซลล์ความจำ (Cell หรือ Memory cell) มีหน้าที่เก็บข้อมูลขาเข้าที่สำคัญเอาไว้ ไม่ให้เลือนหายไปในช่วงระยะเวลาหนึ่ง เป็นการแก้ปัญหาการพึ่งพาระยะยาวซึ่งได้กล่าวไปแล้วในหัวข้อก่อนหน้า สำหรับกลไกการตัดสินใจของเซลล์ความจำว่าจะเลือกเก็บข้อมูลขาเข้าอย่างไรนั้น ขึ้นอยู่กับค่าที่ได้รับจาก “ประตูสัญญาณ” (Gate) จำนวน 3 ประตู ซึ่งรับ s_{t-1} กับ x_t ไปประมวลผลในลักษณะเดียวกันกับโครงข่ายประสาทเทียมแบบวกกลับโดยทั่วไป ประตูสัญญาณเหล่านี้ประกอบด้วย

1) ประตูสัญญาณข้อมูลขาเข้า (Input gate) มีหน้าที่ประเมินความสำคัญของ x_t ซึ่งเข้ามาใหม่
2) ประตูสัญญาณสำหรับลืม (Forget gate) มีหน้าที่ประเมินความสำคัญของข้อมูลที่เก็บในเซลล์ความจำว่ายังควรจำไว้อยู่ หรือสมควรที่จะลืมแล้วแทนที่ด้วยข้อมูลใหม่
3) ประตูสัญญาณข้อมูลผลลัพธ์ (Output gate) มีหน้าที่ประเมินว่าข้อมูลที่เก็บในเซลล์ความจำควรนำมาใช้ในการให้ผลลัพธ์ของโครงข่ายความจำระยะสั้นขนาดยาวแค่ไหนเพียงใด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

กำหนดให้ i คือ ประตูสัญญาณข้อมูลขาเข้า

f คือ ประตูสัญญาณสำหรับลืม

o คือ ประตูสัญญาณข้อมูลผลลัพธ์

c คือ เซลล์ความจำ

σ คือ ฟังก์ชันซิกมอยด์

$\circ$ คือ ผลคูณแบบอาดามาร์ (Hadamard product) ซึ่งเป็นการนำสมาชิกของเมทริกซ์สองตัวที่มีแถวและหลักเดียวกันมาคูณกัน (Element-wise product)

ชุดสมการแสดงการทำงานของโครงข่ายความจำระยะสั้นขนาดยาว แสดงได้ดังนี้

$$i_t = \sigma(W_i s_{t-1} + U_i x_t + b_i) \quad (2.28)$$

$$f_t = \sigma(W_f s_{t-1} + U_f x_t + b_f) \quad (2.29)$$

$$o_t = \sigma(W_o s_{t-1} + U_o x_t + b_o) \quad (2.30)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh(W_c s_{t-1} + U_c x_f + b_c) \quad (2.31)$$

$$s_t = o_t \circ \sigma(c_t) \quad (2.32)$$

2.7.3 โครงข่ายประตูวกกลับ (Gated Recurrent Unit : GRU) [22]

โครงข่ายประตูวกกลับ พัฒนาขึ้นเมื่อค.ศ.2014 มีลักษณะคล้ายโครงข่ายความจำระยะสั้นขนาดยาว [23] แต่มีขนาดเล็กลง เพื่อประหยัดเนื้อที่หน่วยความจำของเครื่องคอมพิวเตอร์และทำให้โครงข่ายทำงานเร็วขึ้น ทั้งนี้ โครงข่ายประตูวกกลับจะยุบประตูสัญญาณข้อมูลขาเข้า และประตูสัญญาณข้อมูลผลลัพธ์รวมกันให้กลายเป็น ประตูสัญญาณอัพเดท (Update gate) และยุบเซลล์ความจำเข้ากับชั้นซ่อน s_t ส่งผลให้เหลือประตูสัญญาณเพียง 2 ประตู และไม่มีเซลล์ความจำแยกออกมาต่างหาก อนึ่ง สำหรับประตูสัญญาณสำหรับลืมนั้น ในโครงข่ายประตูวกกลับจะเรียกว่า ประตูสัญญาณรีเซ็ต (Reset gate) [23]

กำหนดให้ z คือ ประตูสัญญาณอัพเดท

r คือ ประตูสัญญาณรีเซ็ต

ชุดสมการแสดงการทำงานของโครงข่ายประตูวกกลับ แสดงได้ดังนี้

Chulalongkorn University

$$z_t = \sigma(W_z s_{t-1} + U_z x_t + b_z) \quad (2.33)$$

$$r_t = \sigma(W_r s_{t-1} + U_r x_t + b_r) \quad (2.34)$$

$$\tilde{s}_t = tanh(W_s(r_t \circ s_{t-1}) + U_s x_t + b_s) \quad (2.35)$$

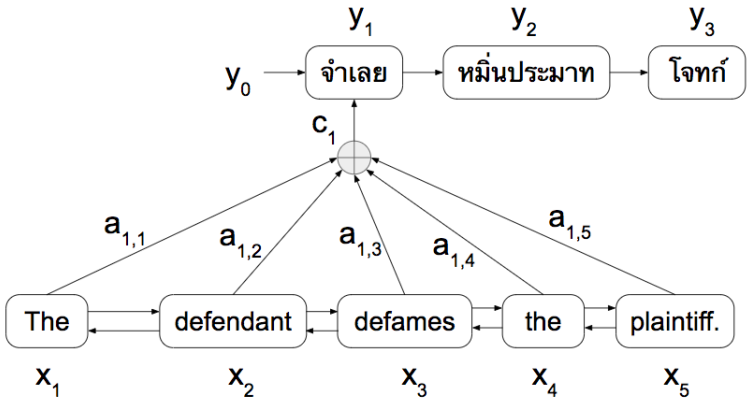
$$s_t = z_t \circ s_{t-1} + (1 - z_t) \circ \tilde{s}_t \quad (2.36)$$

2.8 กลไกจุดสนใจ (Attention Mechanism)

กลไกจุดสนใจได้รับการพัฒนาขึ้นครั้งแรกเมื่อค.ศ.2014 [24] เพื่อใช้ในงานแปลภาษาผ่านเครื่องด้วยโครงข่ายประสาทเทียม (Neural machine translation) ต่อมาได้รับการนำไปใช้ในงานด้านต่าง ๆ ที่ใช้โครงข่ายประสาทเทียมเพื่อประมวลผลข้อมูลแบบลำดับต่อลำดับ (Sequence-to-sequence model) เช่น การตอบคำถาม (Question answering) [25, 26] อนึ่ง โครงข่ายประสาทเทียมในหัวข้อนี้ คือ โครงข่ายประสาทเทียมแบบวกกลับ

หลักการของกลไกนี้จะสร้างค่าความสนใจ (Attention) ให้กับสมาชิกแต่ละตัวในลำดับข้อมูล แล้วนำค่าที่ได้คูณกลับไปหาสมาชิกตัวนั้น ๆ เพื่อสร้างเวกเตอร์ผลลัพธ์ที่จะนำไปใช้ในโครงข่ายประสาทเทียมต่อไป ค่าความสนใจนี้จึงเปรียบเสมือนค่าน้ำหนักที่ถ่วงข้อมูลแต่ละตัวในลำดับข้อมูลนั้น ถ้าข้อมูลไหนควรสนใจมาก ค่าความสนใจก็จะมากด้วย

ตัวอย่างการประยุกต์ใช้กลไกจุดสนใจพิจารณาได้จากงานแปลข้อความภาษาอังกฤษเป็นภาษาไทย ซึ่งสามารถพิจารณารูปที่ 2.9 ประกอบได้ ทั้งนี้ การแปลข้อความภาษาอังกฤษ “The defendant defames the plaintiff.”เป็นภาษาไทยว่า “จำเลยหมิ่นประมาทโจทก์” นั้น จะใช้โครงข่ายประสาทเทียม 2 โครงข่ายทำงานในลักษณะตัวเข้ารหัส - ตัวถอดรหัส (Encoder - Decoder) โดยนำเวกเตอร์คำภาษาอังกฤษทั้งหมดไปประมวลผลผ่านตัวเข้ารหัสก่อน ส่งผลให้ได้ผลลัพธ์จากชั้นซ่อนของตัวเข้ารหัสจากคำต่าง ๆ ในประโยคภาษาอังกฤษ แทนด้วย x_1 ถึง x_5 ต่อมาจึงสร้างเวกเตอร์คำภาษาไทยผ่านตัวถอดรหัส



รูปที่ 2.9 *ตัวอย่างการประยุกต์ใช้กลไกจุดสนใจตาม* [24] *เพื่อแปลข้อความภาษาอังกฤษ “The defendant defames the plaintiff.” เป็นภาษาไทยว่า “จำเลยหมิ่นประมาทโจทก์” ด้วยโครงข่ายประสาทเทียมแบบวกกลับ*

ในการสร้างเวกเตอร์คำภาษาไทยนั้น ถ้าเราพิจารณาคำแรก ได้แก่ “จำเลย” (y_1) แล้ว เราจะนำกลไกจุดสนใจมาใช้ในช่วงนี้ โดยหาค่าความสนใจเมื่อเทียบคำว่า “จำเลย” กับคำภาษาอังกฤษในประโยคต้นทางแต่ละคำ ทั้งนี้ ค่าความสนใจระหว่างคำว่า “จำเลย” เทียบกับคำว่า “The” แทนด้วย $a_{1,1}$ ค่าความสนใจระหว่างคำว่า “จำเลย” เทียบกับคำว่า “defendant” แทนด้วย $a_{1,2}$ ฯลฯ สุดท้ายแล้ว เราจะได้ค่าความสนใจเท่ากับจำนวนคำ

ภาษาอังกฤษทั้งหมดในประโยคต้นทาง แทนด้วย $a_{1,1}$ ถึง $a_{1,5}$ โดยทั้งหมดจะนำไปคูณกลับเข้าหา x_1 ถึง x_5 แล้วนำผลลัพธ์แต่ละเวกเตอร์ที่ได้บวกรวมกันเกิดเป็นเวกเตอร์บริบท (c_1) ซึ่งจะนำไปประมวลผลร่วมกับค่าสถานะจากชั้นซ่อนของตัวถอดรหัสในคำก่อน (y_0) เกิดเป็นเวกเตอร์ของคำว่า “จำเลย” ต่อไป อนึ่ง วิธีนี้จะใช้กับคำว่า “หมิ่นประมาท” (y_2) และ “โจทก์” (y_3) เช่นกัน

กำหนดให้ y_i คือ เวกเตอร์คำลำดับที่ i ของประโยคปลายทาง

RNN คือ โครงข่ายประสาทเทียมที่เลือกใช้ เช่น LSTM หรือ GRU

c_i คือ เวกเตอร์บริบทของคำลำดับที่ i จากประโยคปลายทาง

m คือ จำนวนคำของประโยคต้นทาง

x_j คือ เวกเตอร์คำลำดับที่ j ของประโยคต้นทางที่ได้จากชั้นซ่อนของ RNN

$a_{i,j}$ คือ ค่าความสนใจของคำลำดับที่ i จากประโยคปลายทางเมื่อเทียบกับคำลำดับที่ j ของประโยคต้นทาง

att คือ ฟังก์ชันสำหรับคำนวณค่าความสนใจ

สมการแสดงการสร้างเวกเตอร์คำแต่ละตัวของประโยคปลายทาง และการคำนวณค่าความสนใจระหว่างคำสองคำ แสดงได้ดังนี้

$$y_i = RNN(y_{i-1}, c_i) \tag{2.37}$$

$$c_i = \sum_{j=1}^{m} a_{i,j} x_j \tag{2.38}$$

$$a_{i,j} = softmax(att(y_i, x_j)) \tag{2.39}$$

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จากสมการ (2.39) ซึ่งเป็นสมการคำนวณค่าความสนใจนั้น ฟังก์ชันคำนวณค่าความสนใจ att มีหลายรูปแบบด้วยกัน หากกำหนดให้ T เป็นสัญลักษณ์ของการสลับเปลี่ยนในเมทริกซ์ (Transpose)

W_x คือ เมทริกซ์ค่าน้ำหนักของลำดับข้อความ x

W_y คือ เมทริกซ์ค่าน้ำหนักของลำดับข้อความ y

W_a คือ เมทริกซ์ค่าน้ำหนักกลางสำหรับลำดับข้อความสองลำดับ

v_a คือ เวกเตอร์ของค่าน้ำหนักที่ใช้หาค่าความสนใจ

ตัวอย่างฟังก์ชันคำนวณค่าความสนใจ มีดังนี้

1) ค่าความสนใจแบบบวก (Additive attention) [24] ใช้โครงข่ายประสาทเทียมแบบไปข้างหน้าดังที่เคยกล่าวมาแล้วในหัวข้อที่ 2.5.2 นำมาคำนวณหาค่าความสนใจ พิจารณาได้จากสมการต่อไปนี้

$$att(y_i, x_j) = v_a^{\mathrm{T}} \cdot \tanh(W_y y_i + W_x x_j) \quad (2.40)$$

2) ค่าความสนใจแบบคูณ (Multiplicative attention) จะใช้เมทริกซ์ค่าน้ำหนักกลางเพียงตัวเดียวในการคำนวณความสนใจ มีข้อดีกว่าค่าความสนใจแบบบวกในเชิงความเร็วการประมวลผล เพราะใช้ค่าน้ำหนักเพียงชุดเดียว พิจารณาได้จากสมการต่อไปนี้

$$att(y_i, x_j) = y_i^{\mathrm{T}} W_a x_j \quad (2.41)$$

3) ค่าความสนใจในตัวเอง (Self-attention) เป็นค่าความสนใจที่มองว่าลำดับข้อมูลสองลำดับ**คือลำดับเดียวกัน** ทำให้ค่าความสนใจชนิดนี้พิจารณาลำดับข้อมูลเพียงลำดับเดียวเท่านั้น สำหรับวิธีการคำนวณจะใช้รูปแบบเดียวกันกับสมการ (2.40) และ (2.41) ขึ้นอยู่กับการเลือกใช้ เพียงแต่ $y_i = x_j$ ดังนั้น จึงอาจเขียนรูปแบบสมการใหม่ได้ โดยขอเลือกใช้ x_j เป็นตัวแทนลำดับข้อความว่า

$$selfAtt(x_j) = v_a^{\mathrm{T}} \cdot \tanh(W_x x_j) \quad (2.42)$$

$$selfAtt(x_j) = x_j^{\mathrm{T}} W_x x_j \quad (2.43)$$

2.9 การวัดประสิทธิภาพการทำงาน (Performance evaluation)

การวัดประสิทธิภาพของแบบจำลองจะใช้ตัวชี้วัดเป็นค่าเอฟวัน ($F1$) อย่างไรก็ดี เนื่องจากค่าเอฟวันมีที่มาจากการหาค่าเฉลี่ยฮาร์มอนิกของค่าความเที่ยงตรง (Precision : P) และค่าความถูกต้อง (Recall : R) ซึ่งทั้งสองล้วนเป็นตัวชี้วัดเช่นเดียวกัน ดังนั้น จะขอแสดงสมการคำนวณตัวชี้วัด 3 ค่า โดยพิจารณาจากการทำนายผลลัพธ์แบบสองคำตอบ ซึ่งมีคำตอบได้ 2 รูปแบบเท่านั้น คือ 1 กับ 0

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

กำหนดข้อมูลการนับดังต่อไปนี้

- TP เป็นจำนวนคำตอบที่ทำนายผลลัพธ์ได้ 1 และคำเฉลยคือ 1 (True Positive)
- FP เป็นจำนวนคำตอบที่ทำนายผลลัพธ์ได้ 1 แต่คำเฉลยเป็น 0 (False Positive)
- TN เป็นจำนวนคำตอบที่ทำนายผลลัพธ์ได้ 0 และคำเฉลยคือ 0 (True Negative)
- FN เป็นจำนวนคำตอบที่ทำนายผลลัพธ์ได้ 0 แต่คำเฉลยเป็น 1 (False Negative)

ตัวชี้วัดทั้งหมดคำนวณได้จากข้อมูลการนับข้างต้น ดังสมการต่อไปนี้

$$P = \frac{TP}{TP + FP} \quad (2.44)$$

$$R = \frac{TP}{TP + FN} \quad (2.45)$$

$$F1 = \frac{2PR}{P + R} \quad (2.46)$$

เนื่องด้วยนิยามของ TP สนใจการทำนายผลลัพธ์ที่มีค่าเป็น 1 ดังนั้น การคำนวณค่าความเที่ยงตรง และค่าความถูกต้องตามสมการ (2.44) และ (2.45) จึงให้ผลลัพธ์ที่พิจารณาจากการทำนายคำตอบเป็น 1 เพียงฝั่งเดียว **ส่งผลให้ค่าเอฟวันที่ได้มาจากการทำนายคำตอบที่เป็น 1 เท่านั้น** ดังนั้น เพื่อการวัดประสิทธิภาพของแบบจำลองในภาพรวม **เราจะหาค่าเอฟวันเฉลี่ยจากทั้งสองคำตอบ** กล่าวคือ หาค่าเอฟวัน เมื่อสนใจผลลัพธ์เป็น 1 และ 0 ตามลำดับ ก่อนนำมาเฉลี่ยกัน ทั้งนี้ การหาค่าเฉลี่ยข้างต้นมี 2 ลักษณะ คือ

1) ค่าเฉลี่ยมหภาค (Macro-average) เป็นการนำค่าตัวชี้วัดแต่ละตัวมาเฉลี่ยตามจำนวนรูปแบบคำตอบ

กำหนดให้ $F1_1$ แทนค่าเอฟวันที่สนใจการทำนายคำตอบเป็น 1

$F1_0$ แทนค่าเอฟวันที่สนใจการทำนายคำตอบเป็น 0

ค่าเฉลี่ยมหภาคของเอฟวัน แสดงได้ตามสมการดังต่อไปนี้

$$F1_{Macro} = \frac{F1_1 + F1_0}{2} \quad (2.47)$$

2) ค่าเฉลี่ยจุลภาค (Micro-average) เป็นการนำค่าข้อมูลการนับจากทุกรูปแบบคำตอบมารวมกัน แล้วนำผลรวมนั้นไปหาค่าตัวชี้วัดโดยตรง ดังนั้น เราจะพิจารณาสมการ (2.44) (2.45) และ (2.46) ใหม่

กำหนดให้ C แทนจำนวนรูปแบบคำตอบทั้งหมดที่สนใจ

M_C แทนค่าข้อมูลการนับ ณ รูปแบบคำตอบที่สนใจ เช่น ถ้า $C = 3$ ประกอบด้วย 0, 1, 2 แล้ว TP จะมีได้ 3 ค่า โดยเป็น TP เมื่อสนใจคำตอบ 0, 1, 2 ตามลำดับ นั่นเอง

ค่าเฉลี่ยจุลภาคของเอฟวัน แสดงได้ตามสมการดังต่อไปนี้

$$P' = \frac{\sum_C TP_C}{\sum_C TP_C + \sum FP_C} \quad (2.48)$$

$$R' = \frac{\sum_C TP_C}{\sum_C TP_C + \sum_C FN_C} \quad (2.49)$$

$$F1_{Micro} = \frac{2P'R'}{P'+R'} \quad (2.50)$$

อย่างไรก็ดี หากพิจารณาเฉพาะกรณีการแบ่งประเภทสองคำตอบ เราจะพบว่า TP_0 ก็คือ TN และ FP_0 ก็คือ FN นั่นเอง ดังนั้น จากสมการ (2.48) และ (2.49) จะแทนค่าเป็น

$$P' = \frac{(TP+TN)}{(TP+TN)+(FP+FN)} \quad (2.51)$$

$$R' = \frac{(TP+TN)}{(TP+TN)+(FN+FP)} \quad (2.52)$$

ณ จุดนี้ จะเห็นได้ว่าค่าความเที่ยงตรงและค่าความถูกต้องที่คำนวณได้มีค่าเท่ากัน และตรงกับนิยามของค่าความแม่นยำ (Accuracy : Acc) ดังนั้น **ในกรณีการแบ่งประเภทสองคำตอบ ค่าเฉลี่ยจุลภาคของเอฟวันก็คือค่าความแม่นยำนั่นเอง** ส่งผลให้ ค่าเฉลี่ยจุลภาคของเอฟวัน สามารถเขียนได้อีกรูปแบบหนึ่งดังนี้

$$F1_{Micro} = Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.53)$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 3
งานวิจัยที่เกี่ยวข้อง

3.1 งานวิจัยเกี่ยวกับการทำนายผลคำพิพากษาของศาลในต่างประเทศ

3.1.1 ผลงานวิจัยโดย Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preotiuc-Pietro และ Vasileios Lampos (ค.ศ.2016) [27]

ผลงานวิจัยชิ้นนี้เป็นการพัฒนาแบบจำลองทำนายผลคำพิพากษาของศาลสิทธิมนุษยชนแห่งยุโรป (The European Court of Human Rights : ECtHR) เมื่อค.ศ.2016 เพื่อเป็นเครื่องมือสำหรับนักกฎหมายและผู้พิพากษาในการค้นหาบรรทัดฐานคำตัดสินในคดีต่าง ๆ ของศาลที่เกี่ยวข้องกับการละเมิดอนุสัญญายุโรปว่าด้วยสิทธิมนุษยชน หรืออนุสัญญาอีซีเอชอาร์ (The European Convention of Human Rights : ECHR) โดยแบบจำลองนี้จะให้ผลลัพธ์ว่าข้อเท็จจริงตามคำพิพากษานั้นเป็นการละเมิดอนุสัญญาอีซีเอชอาร์หรือไม่ ซึ่งคำตอบจะมีได้ 2 แบบเท่านั้น คือ ละเมิด กับ ไม่ละเมิด ด้วยลักษณะคำตอบนี้ ทำให้แบบจำลองเข้าข่ายเป็นตัวจำแนกประเภทแบบสองคำตอบนั่นเอง ทั้งนี้ คณะผู้วิจัยได้ใช้ข้อมูลคำพิพากษาศาลสิทธิมนุษยชนแห่งยุโรปที่ตัดสินตามอนุสัญญาอีซีเอชอาร์จำนวน 584 คดี มาเป็นข้อมูลในการสร้างและทดสอบแบบจำลอง โดยแบ่งเป็นคำตัดสินตามมาตรา 3 จำนวน 250 คดี มาตรา 6 จำนวน 80 คดี และมาตรา 8 จำนวน 254 คดี นอกจากนี้กลุ่มคำพิพากษาในแต่ละมาตราจะแบ่งสัดส่วนตามผลคำพิพากษาได้เท่ากัน คือ มีทั้งที่ตัดสินว่า ละเมิด และ ไม่ละเมิดอย่างละครึ่งด้วย

แบบจำลองจะรับข้อมูลขาเข้าเป็นข้อความในคำพิพากษา จากนั้นจะนำข้อความดังกล่าวไปสร้างเอ็นแกรม ที่มีขนาดของ N ตั้งแต่ 1 ถึง 4 ก่อนจะคัดเลือกเอ็นแกรมที่มีความถี่สูงสุด 2,000 ค่า มาใช้เป็นตัวแทนของข้อความในคำพิพากษาดังกล่าว เมื่อได้ตัวแทนของข้อความแล้ว จึงนำไปเข้าอัลกอริทึม ซัพพอร์ตเวกเตอร์แมชชีน หรือ เอสวีเอ็ม (Support Vector Machine : SVM) เพื่อให้ได้ผลลัพธ์เป็นผลคำพิพากษาว่าคดีนั้นละเมิดอนุสัญญาอีซีเอชอาร์หรือไม่ ในท้ายที่สุด อย่างไรก็ดี นอกเหนือจากการแทนข้อความแบบเอ็นแกรมแล้ว คณะผู้วิจัยได้ทดลองแทนข้อความอีกรูปแบบหนึ่ง โดยหลังจากได้เอ็นแกรมที่มีความถี่สูงสุด 2,000 ค่าของคำพิพากษาแต่ละฉบับแล้ว จะนำเอ็นแกรมของคำพิพากษาที่ตัดสินตามมาตราเดียวกันไปสร้างเมทริกซ์ความคล้าย (Similarity matrix) ระหว่าง เอ็นแกรมของคำพิพากษาแต่ละฉบับ แล้วนำเมทริกซ์ความคล้ายนี้ไปแบ่งกลุ่มแบบสเปกตรัม (Spectral clustering) เพื่อจัดกลุ่มเอ็นแกรมที่มีความหมายใกล้เคียงกัน หรืออีกนัยหนึ่ง คือ จัดกลุ่มคำพิพากษาที่มีลักษณะคดีคล้ายกันด้วย ผลลัพธ์ที่ได้จากการแบ่งกลุ่มแบบสเปกตรัม คือ เซตของเอ็นแกรมที่อยู่ในกลุ่มเดียวกัน ซึ่งคณะผู้วิจัยจะเรียกว่า หัวข้อ (Topic) และนำมาใช้เป็นตัวแทนของข้อความที่จะนำไปเข้าเอสวีเอ็มต่อไปด้วย

สำหรับการทดสอบแบบจำลองนั้น จะแบ่งชุดข้อมูลสอนและข้อมูลทดสอบด้วยวิธีการตรวจสอบความสมเหตุสมผลแบบข้ามส่วน K ส่วน (K-fold Cross-validation) โดยกำหนด K เป็น 10 หลังจากนั้นในแต่ละรอบจะนำข้อความในคำพิพากษาส่วนต่าง ๆ ประกอบด้วย กระบวนพิจารณา หรือรายละเอียดการดำเนินคดี (Procedure), สถานการณ์ในคดี (Circumstances), กฎหมายภายในประเทศที่เกี่ยวข้องกับคดี (Relevant law)

และ ข้อกฎหมายในคดีตามอนุสัญญาอีซีเอชอาร์ (Law) มานำเข้าแบบจำลองแล้ววัดค่าความแม่นยำ นอกจากนี้ คณะผู้วิจัยยังรวมข้อมูล สถานการณ์ในคดี และ กฎหมายภายในประเทศฯ เข้าด้วยกันเรียกว่า ข้อเท็จจริง (Facts) และรวมข้อความทั้งหมด คือ กระบวนพิจารณา สถานการณ์ในคดี กฎหมายภายในประเทศฯ และ ข้อกฎหมาย เข้าด้วยกัน เรียกว่า ข้อความทั้งหมด (Full) ก่อนจะนำ ข้อเท็จจริง กับ ข้อความทั้งหมด ไปเข้าแบบจำลองด้วย ยิ่งไปกว่านั้น คณะผู้วิจัย ได้ทดลองเปลี่ยนวิธีแทนข้อความจากเอ็นแกรมปกติ มาเป็นการแทนข้อความด้วยหัวข้อที่สร้างขึ้น ตลอดจนการใช้ทั้งเอ็นแกรมและหัวข้อประกอบกันด้วย ผลการทดสอบพบว่า

- ในการทำนายผลคำตัดสินคดีตามมาตรา 3 แห่งอนุสัญญาอีซีเอชอาร์ พบว่า การแทนข้อความด้วยหัวข้อ ทำให้แบบจำลองมีค่าความแม่นยำเฉลี่ยสูงสุดอยู่ที่ 78%
- ในการทำนายผลคำตัดสินคดีตามมาตรา 6 และมาตรา 8 แห่งอนุสัญญาอีซีเอชอาร์ พบว่า การแทนข้อความด้วย หัวข้อ ผสมกับเอ็นแกรมของสถานการณ์ในคดี ทำให้แบบจำลองมีค่าความแม่นยำเฉลี่ยสูงสุดอยู่ที่ 84% และ 78% ตามลำดับ
- ในภาพรวมของการทำนายผลคำตัดสินคดีทั้ง 3 มาตรา เมื่อนำค่าความแม่นยำเฉลี่ยของแต่ละมาตรามาเฉลี่ยกัน พบว่าการแทนข้อความด้วย หัวข้อ ผสมกับเอ็นแกรมของสถานการณ์ในคดี ทำให้แบบจำลองมีค่าความแม่นยำเฉลี่ยสูงสุดอยู่ที่ 79%

ดังนั้น คณะผู้วิจัยจึงเห็นว่า ข้อเท็จจริงในคดี (ในที่นี้ คือ สถานการณ์ในคดี) คือปัจจัยที่สำคัญที่สุดในการทำนายผลคำพิพากษา สอดคล้องกับการพิจารณาคดีของศาลที่ต้องปรับข้อกฎหมายให้สอดคล้องกับข้อเท็จจริงตลอดจน หัวข้อของคำพิพากษา นับเป็นคุณลักษณะหนึ่งที่สำคัญต่อการทำนายหรือการแบ่งประเภทข้อมูลด้วย อนึ่ง ในการทดสอบแบบจำลองครั้งนี้ คณะผู้วิจัย ไม่ได้ ทดลองเปรียบเทียบเอสวีเอ็ม ที่ใช้อยู่กับอัลกอริทึมอื่นแต่อย่างใด

สำหรับข้อแตกต่างระหว่างผลงานวิจัยค.ศ.2016 ชิ้นนี้กับแบบจำลองที่ผู้วิจัยพัฒนาขึ้น มีดังนี้

1) ผลงานวิจัยชิ้นนี้ใช้ชุดข้อมูลเป็นคำพิพากษาภาษาอังกฤษของศาลสิทธิมนุษยชนแห่งยุโรป เฉพาะที่ตัดสินตามมาตรา 3 มาตรา 6 และมาตรา 8 แห่งอนุสัญญาอีซีเอชอาร์ จำนวนรวม 254 ฉบับ ขณะที่แบบจำลองที่ผู้วิจัยพัฒนาขึ้น ใช้ชุดข้อมูลเป็นคำพิพากษาศาลฎีกาฉบับย่อในคดีอาญาซึ่งเป็นภาษาไทย และตัดสินคดีภายใต้ประมวลกฎหมายอาญาของไทย ในคดีความผิดเกี่ยวกับชีวิตและร่างกาย ชื่อเสียง และทรัพย์ รวมจำนวนได้ 1,000 ฉบับ
2) ผลงานวิจัยชิ้นนี้สร้างตัวแทนข้อความด้วยแบบจำลองเอ็น-แกรม และทำนายคำตัดสินด้วยเอสวีเอ็ม ส่วนแบบจำลองที่ผู้วิจัยพัฒนาขึ้นใช้เทคนิคการเรียนรู้เชิงลึก ซึ่งสามารถสร้างตัวแทนข้อความและทำนายผลคำตัดสินได้ด้วยโครงข่ายประสาทเทียมทั้งแบบวกกลับและแบบไปข้างหน้า

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.1.2 ผลงานวิจัยโดย Daniel Martin Katz, Michael J. Bommarito II และ Josh Blackman (ค.ศ.2017) [28]

ผลงานวิจัยชิ้นนี้เป็นการพัฒนาแบบจำลองทำนายผลคำพิพากษาของศาลสูงแห่งสหรัฐอเมริกา (The Supreme Court of the United States) เมื่อปีค.ศ.2017 เพื่อวิเคราะห์ว่า ศาลสูงสหรัฐอเมริกาจะมีคำพิพากษายืน (Affirm) หรือกลับ (Reverse) คำพิพากษาของศาลล่าง ทั้งในเชิงของคำพิพากษาทั้งคดี และคำตัดสินของผู้พิพากษารายคน (คำพิพากษาทั้งคดีมาจากเสียงข้างมากของคำตัดสินจากผู้พิพากษารายคนทั้งหมด) ทั้งนี้ คณะผู้วิจัยได้ใช้ฐานข้อมูล เอสซีดีบี (SCDB ย่อมาจาก Supreme Court Database) ซึ่งพัฒนาขึ้นโดย Harold J. Spaeth และคณะ [29] เป็นชุดข้อมูลในการวิจัย โดยฐานข้อมูลนี้ได้รวบรวมและสกัดข้อมูลจากคำพิพากษาศาลสูงสหรัฐอเมริกาตั้งแต่ค.ศ.1791 ซึ่งเป็นปีที่ศาลสูงได้ตัดสินคดีแรก จนกระทั่งถึงค.ศ.2016 (ณ เดือนเมษายน ค.ศ.2018) โดยคำพิพากษาแต่ละฉบับจะมีคุณลักษณะอยู่ 53 คุณลักษณะ และหากลงรายละเอียดถึงคำตัดสินของผู้พิพากษารายคน จะมีคุณลักษณะเพิ่มขึ้นอีก 8 คุณลักษณะ รวมทั้งหมดเป็น 61 คุณลักษณะ ส่วนใหญ่เป็นข้อมูลชนิดประเภท (Categorical variable) โดยระบุไว้เป็นตัวเลข มีเพียงส่วนน้อยที่เป็นข้อมูลตัวอักษร เช่น ชื่อคดี หรือชื่อประธานศาลสูงสุดที่ตัดสินคดีนั้น สำหรับข้อมูลที่นำมาใช้ในผลงานวิจัยชิ้นนี้มาจากคำพิพากษาที่ตัดสินระหว่างค.ศ.1816 จนถึงค.ศ.2015 รวม 28,009 ฉบับ และจำแนกเป็นผลคำตัดสินของผู้พิพากษาได้ 243,882 คำตัดสิน

ทั้งนี้ แบบจำลองจะรับข้อมูลขาเข้าเป็นคุณลักษณะบางส่วนในฐานข้อมูลเอสซีดีบี จำนวน 13 คุณลักษณะ ยกตัวอย่างเช่น หมายเลขประจำตัวผู้พิพากษา (JUSTICE ID), ปีที่ตัดสินคดี (TERM), ผู้ยื่นคำร้อง (PETITIONER) ซึ่งอาจเป็นโจทก์หรือจำเลยในคดีนั้นก็ได้, ฝ่ายตรงข้ามผู้ยื่นคำร้อง (RESPONDENT), ศาลชั้นต้นที่พิจารณาคดี (COURT OF ORIGIN), ศาลชั้นอุทธรณ์ที่พิจารณาคดี (SOURCE OF THE CASE), ผลคำตัดสินของศาลล่าง (LOWER COURT DISPOSITION), หัวข้อของคดี (ISSUE), ประเภทคดี (ISSUE AREA) ฯลฯ จากนั้นจะนำคุณลักษณะข้างต้นไปผ่านกระบวนการวิศวกรรมข้อมูลตัวแทน (Feature engineering) เพื่อให้ได้ข้อมูลที่มีลักษณะเป็นสองค่า (Binary) ก่อนนำไปเข้าอัลกอริทึม ป่าสุ่ม (Random forest) จนได้ผลลัพธ์สุดท้ายเป็นคำตัดสินว่า คดีนี้ศาลพิพากษายืนหรือกลับคำตัดสินของศาลล่างนั่นเอง อย่างไรก็ดี เฉพาะการทำนายผลคำตัดสินของผู้พิพากษาเป็นรายบุคคล จะมีผลลัพธ์ได้สามค่า คือ ยืน กลับ หรือ “อย่างอื่น” (Other) เช่น ในกรณีที่ผู้พิพากษาถอนตัวจากคดี (Recusal) หรือในคดีที่ยื่นฟ้องต่อศาลสูงได้โดยตรง ทำให้ไม่มีคดีจากศาลล่างมาเปรียบเทียบว่าศาลสูงพิพากษายืนหรือกลับ ฯลฯ อนึ่ง ในการสอนแบบจำลองนี้ คณะผู้วิจัยระบุว่าจำเป็นอย่างยิ่งที่ต้องพิจารณาปีที่ตัดสินคดีด้วย เพราะข้อมูลสอนจะต้องเป็นคำพิพากษาที่ตัดสิน **ก่อน** ปีที่สนใจจะทำนายด้วยเสมอ เช่น ถ้าต้องการทดสอบแบบจำลองโดยทำนายคำตัดสินคดีในปีค.ศ.2010 จะต้องใช้ข้อมูลสอนเป็นคำพิพากษาก่อนปีค.ศ.2010 ด้วยเสมอ การกำหนดเช่นนี้เป็นผลจากระบบกฎหมายของสหรัฐอเมริกาที่ใช้ระบบคอมมอน ลอว์ (Common law) ซึ่งคำพิพากษาเป็นที่มาของกฎหมายประการหนึ่งด้วยนั่นเอง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สำหรับการทดสอบแบบจำลองนั้น จะทดสอบกับข้อมูลทั้งในระดับผลคำตัดสินของผู้พิพากษารายบุคคล และผลคำตัดสินทั้งคดี โดยเปรียบเทียบกับ แบบจำลองเปรียบเทียบ (Baseline model) 3 แบบ กล่าวคือ

- แบบแรกให้ทำนายผลลัพธ์ว่า กลับคำตัดสินเสมอ
- แบบที่สองให้ทำนายจาก ผลคำตัดสินที่มีจำนวนมากที่สุดในบรรดาผลคำตัดสินทั้งหมดจนถึงช่วงก่อนปีที่สนใจ เช่น ถ้าทำนายคำตัดสินคดีในปีค.ศ.2010 และมีข้อมูลสอนเริ่มต้นตั้งแต่ค.ศ.1815 แล้ว จะพิจารณาข้อมูลสอนตั้งแต่ค.ศ.1815 – 2009 ถ้าให้ผลลัพธ์ว่า กลับคำตัดสิน มากที่สุด แบบจำลองนี้จะให้ผลลัพธ์ว่า กลับคำตัดสิน
- แบบที่สามให้ทำนายจาก ผลคำตัดสินที่มีจำนวนมากที่สุดในบรรดาผลคำตัดสินย้อนหลัง 10 ปี ก่อนปีที่สนใจจะทำนาย ซึ่งถ้าเป็นตัวอย่างก่อนหน้า ก็จะพิจารณาจากข้อมูลสอนตั้งแต่ค.ศ.2000 – 2009 แทนที่จะเป็นค.ศ.1815 นั่นเอง

เหตุที่คณะผู้วิจัยทำแบบจำลองเปรียบเทียบลักษณะนี้ เนื่องจากต้องการให้สอดคล้องกับแนวทางการวิเคราะห์คำพิพากษาของนักกฎหมายสหรัฐอเมริกาโดยทั่วไป ซึ่งมักเห็นว่าคำพิพากษาศาลสูงจะกลับคำพิพากษาของศาลล่างเสมอ สำหรับผลการทดลองนั้น พบว่า

- ในการวัดค่าความแม่นยำนั้น การทำนายผลคำตัดสินเป็นรายคนมีค่าความแม่นยำ 70.2% สูงกว่าแบบจำลองเปรียบเทียบแบบที่สาม ซึ่งได้ค่าความแม่นยำ 66.2% ขณะที่การทำนายผลคำตัดสินเป็นรายคดี เมื่อนำข้อมูลคำพิพากษาตั้งแต่ค.ศ.1816 ถึงเดือนตุลาคม ค.ศ.2014 มาทดสอบพบว่า แบบจำลองมีค่าความแม่นยำ 71.9% สูงกว่าแบบจำลองเปรียบเทียบแบบที่สาม ซึ่งได้ค่าความแม่นยำอยู่ที่ 67.5%
- ในการวัดค่าเอฟวัน (F1) พบว่า การทำนายผลคำตัดสินเป็นรายคน มีค่าเอฟวันเฉลี่ยทุกคำตอบอยู่ที่ 64% มากกว่าแบบจำลองเปรียบเทียบแบบที่สาม ซึ่งได้ค่าเอฟวันเฉลี่ยทุกคำตอบเพียง 48% ขณะที่การทำนายผลคำตัดสินเป็นรายคดี พบว่าแบบจำลองมีค่าเอฟวันเฉลี่ยอยู่ที่ 69% สูงกว่าแบบจำลองเปรียบเทียบแบบที่สาม ซึ่งมีค่าเอฟวันเฉลี่ย 66%

ดังนั้น คณะผู้วิจัยจึงเห็นว่าแบบจำลองที่พัฒนาขึ้นสามารถทำนายผลคำตัดสินของศาลสูงสหรัฐอเมริกาได้อย่างมีประสิทธิภาพสอดคล้องกับการวินิจฉัยคดีที่เกิดขึ้นจริง อย่างไรก็ดี มีข้อสังเกตว่า ข้อมูลที่นำใช้ในการสร้างแบบจำลองนี้ ไม่ได้ใช้ข้อมูลที่เป็นตัวอักษรเลย เพราะข้อมูลได้รับการจัดวางประเภทมาเรียบร้อยแล้ว นอกจากนี้ แม้คณะผู้วิจัยจะเลือกใช้อัลกอริทึม ป่าสุ่ม โดยให้เหตุผลประการหนึ่งว่ามีประสิทธิภาพดีกว่าอัลกอริทึมอื่น เช่น เอสวีเอ็มหรือโครงข่ายประสาทเทียมแบบไปข้างหน้าก็ตาม แต่คณะผู้วิจัย ไม่ได้ แสดงผลการวัดค่าความแม่นยำหรือค่าเอฟวัน ระหว่างอัลกอริทึมป่าสุ่ม กับอัลกอริทึมอื่นที่ได้กล่าวมานี้แต่อย่างใด

อย่างไรก็ดี ผลงานวิจัยค.ศ.2017 ชิ้นนี้มีข้อแตกต่างจากแบบจำลองที่ผู้วิจัยพัฒนาขึ้น กล่าวคือ

1) ผลงานวิจัยชิ้นนี้ใช้ฐานข้อมูล เอสซีดีบี ซึ่งสกัดข้อมูลจากคำพิพากษาศาลสูงสหรัฐอเมริกาตั้งแต่ค.ศ.1791 จนถึงค.ศ.2016 แต่คุณลักษณะส่วนใหญ่เป็นข้อมูลชนิดประเภท โดยระบุไว้เป็นตัวเลข มีเพียงส่วนน้อยที่เป็นตัวอักษร ทำให้ลักษณะของชุดข้อมูลนี้ มุ่งเน้นให้นำไปใช้กับเทคนิคการเรียนรู้ของเครื่องได้โดยตรง ส่วนแบบจำลองที่ผู้วิจัยพัฒนาขึ้น ใช้ชุดข้อมูลเป็นคำพิพากษาศาลฎีกาฉบับย่อในคดีอาญาซึ่งเป็นข้อความภาษาไทย โดยรวบรวมตั้งแต่พ.ศ.2501 จนถึงพ.ศ.2559 (ค.ศ.1958 - 2016) ข้อความคำพิพากษาและข้อกฎหมายจากชุดข้อมูลดังกล่าวต้องแปลงให้อยู่ในรูปตัวแทนข้อความก่อนนำไปประมวลผลด้วย

2) ผลงานวิจัยชิ้นนี้ให้ผลลัพธ์การทำนายของแบบจำลองว่าศาลสูงสหรัฐอเมริกาจะพิพากษายืนหรือกลับคำพิพากษาของศาลล่าง แตกต่างจากแบบจำลองที่ผู้วิจัยพัฒนาขึ้น ซึ่งจะให้ผลลัพธ์การทำนายว่า จำเลยมีความผิดตามข้อกฎหมายที่ระบุมาหรือไม่

3.2 งานวิจัยทางคอมพิวเตอร์เกี่ยวกับคำพิพากษาศาลฎีกาไทย

3.2.1 ผลงานวิจัยโดย โษฑศ์รัตต ธรรมบุษดี และ อุดม ศิลปอาชา (ค.ศ.2008) [30]

ผลงานวิจัยชิ้นนี้เป็นการพัฒนาระบบการให้เหตุผลในคำตัดสินคดีอาญาโดยใช้เทคนิคเหมืองข้อมูล เมื่อค.ศ.2008 ทั้งนี้ ข้อมูลทั้งหมดที่นำมาใช้ในการสอนและทดสอบระบบมาจากชุดข้อมูล ทีซีเอ็มแอล (TCXML ย่อมาจาก Thai Court XML) ที่รวบรวมข้อมูลคำพิพากษาศาลฎีกาในคดีอาญา ได้แก่ ข้อหา, ข้อเท็จจริงในชั้นพนักงานสอบสวนและพนักงานอัยการ, ข้อเท็จจริงที่ศาลพิจารณาได้ความ, การวินิจฉัยและผลคำตัดสิน ก่อนนำมาจัดให้อยู่ในรูปแบบภาษาเอ็กซ์เอ็มแอล (XML ย่อมาจาก Extensible Markup Language)

ระบบนี้แบ่งการทำงานเป็น 3 ส่วน คือ

1) ส่วนสกัดองค์ประกอบความผิด (Crime parser) ทำหน้าที่สกัดองค์ประกอบความผิดจากข้อความข้อเท็จจริงโดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision tree) สร้างต้นไม้ตามจำนวนองค์ประกอบความผิดทางอาญา ผลลัพธ์ที่ได้ คือ เซตของผลลัพธ์แบบสองคำตอบตามจำนวนต้นไม้ว่าข้อเท็จจริงเข้าองค์ประกอบความผิดแต่ละข้อหรือไม่
2) ส่วนเชื่อมโยงฐานความผิด (Charges mapper) ทำหน้าที่เชื่อมโยงเซตของผลลัพธ์จากส่วนสกัดองค์ประกอบความผิดเข้ากับตัวบทกฎหมายที่เกี่ยวข้องโดยใช้ตารางค้นหา ผลลัพธ์ที่ได้ คือ รหัสของข้อกฎหมายที่เกี่ยวข้องกับข้อเท็จจริงนั้น
3) ส่วนกำหนดโทษ (Sentences discover) ทำหน้าที่รับผลลัพธ์จากสองส่วนแรกไปประมวลผล เพื่อระบุโทษที่จะได้รับ พร้อมทั้งตรวจสอบด้วยว่าการกระทำดังกล่าวมีเหตุยกเว้นความผิด หรือเหตุยกเว้นโทษหรือไม่ ตลอดจนเป็นการกระทำความผิดกรรมเดียวผิดต่อกฎหมายหลายบท หรือ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

หลายกรรมต่างกัน โดยเทคนิคที่ใช้เทคนิคในส่วนนี้ คือ การถดถอยแบบคลุมเครือ (Fuzzy regression) ผลลัพธ์ที่ได้ คือ โทษที่จะได้รับโดยระบุเป็นช่วง

อนึ่ง ผลงานวิจัยชิ้นนี้ ไม่ได้ แสดงข้อมูลการทดลองเพื่อทดสอบประสิทธิภาพการทำงานของระบบแต่อย่างใด อย่างไรก็ดี ผลงานวิจัยค.ศ.2008 ชิ้นนี้แตกต่างจากแบบจำลองที่ผู้วิจัยพัฒนาขึ้น กล่าวคือ

1) แบบจำลองที่ผู้วิจัยพัฒนาขึ้นจะใช้ชุดข้อมูลที่สร้างขึ้นมาใหม่เอง เนื่องจากไม่พบการเผยแพร่ชุดข้อมูลทีซีเอ็มแอล สำหรับนำไปทดลองตามได้แต่ประการใด

2) แบบจำลองที่ผู้วิจัยพัฒนาขึ้นจะใช้เทคนิคการเรียนรู้เชิงลึกเพื่อประมวลข้อความข้อเท็จจริงและข้อกฎหมายให้กลายเป็นข้อมูลตัวแทน (ตัวแทนข้อความ) ก่อนนำไปสู่ผลลัพธ์สุดท้าย ซึ่งแตกต่างจากผลงานวิจัยชิ้นนี้ที่ใช้กระบวนการสกัดข้อมูลตัวแทนให้อยู่ในรูปตัวเลขด้วยต้นไม้ตัดสินใจ ก่อนนำไปหาผลลัพธ์สุดท้าย

3) แบบจำลองที่ผู้วิจัยพัฒนาขึ้นสามารถสร้างข้อมูลตัวแทนสำหรับประมวลผลหาคำตอบได้ โดยไม่ต้องแยกกระบวนการสกัดข้อมูลตัวแทนเป็นส่วนสกัดองค์ประกอบความผิด และส่วนเชื่อมโยงฐานความผิด เหมือนอย่างผลงานวิจัยชิ้นนี้

3.2.2 ผลงานวิจัยโดย โษฑศ์รัตต ธรรมบุษดี, บัณฑิต วรรธนาภา, Jonathan H. Chan และ อุดม ศิลปอาชา (ค.ศ.2014) [31]

ผลงานวิจัยชิ้นนี้เป็นการพัฒนาตัวจำแนกประเภทสองขั้นตอนเพื่อระบุฐานความผิดและการลงโทษตามกฎหมายอาญา เมื่อค.ศ.2014 เพื่อทำนายคุณลักษณะทางกฎหมายจากข้อเท็จจริงในคดีความผิดเกี่ยวกับชีวิตและร่างกายตามมาตรา 288 ถึงมาตรา 297 แห่งประมวลกฎหมายอาญา นำไปสู่การหาฐานความผิดและบทลงโทษที่ผู้กระทำความผิดจะได้รับ เช่น ถ้ามีข้อเท็จจริงคดีฆ่าผู้อื่นโดยเจตนา ผลลัพธ์ที่ได้ คือ ฐานความผิดและระวางโทษตามมาตรา 288

ตัวจำแนกประเภทที่สร้างขึ้นจะใช้ข้อมูลสอนและทดสอบจากชุดข้อมูลทีซีเอ็มแอล ตามหัวข้อที่ 3.2.1 โดยตั้งต้นจากคุณลักษณะของข้อเท็จจริง (Fact-level attribute) ที่สกัดจากคำพิพากษาแต่ละฉบับไว้แล้วรวมทั้งสิ้น 99 คุณลักษณะ เช่น อาวุธที่ผู้กระทำความผิดใช้, ขนาดของอาวุธ, ขนาดบาดแผล, อวัยวะที่ถูกกระทำ, ลักษณะการกระทำ ฯลฯ จากนั้นจะนำเข้าสู่ตัวจำแนกประเภทสองขั้นตอน แต่ละขั้นตอนจะทำนายคุณลักษณะดังนี้

1) การทำนายคุณลักษณะของคดี (Case-level attribute) โดยใช้คุณลักษณะของข้อเท็จจริงที่กล่าวมาแล้วข้างต้น คุณลักษณะของคดีมีทั้งหมด 25 คุณลักษณะ เช่น เจตนาของการกระทำ, การไตร่ตรองไว้ก่อน, ความรุนแรงของการกระทำ ฯลฯ ทั้งนี้ มี 13 คุณลักษณะด้วยกันที่ต้องเรียนรู้โดยใช้เทคนิคพีซีเอ (PCA ย่อมาจาก Principal Component Analysis) เพื่อลดมิติ และใช้โครงข่าย

ประสาทเทียมแบบไปข้างหน้าเพื่อจำแนกคุณลักษณะออกมา ส่วนอีก 12 คุณลักษณะจะใช้การเขียนกฎซึ่งไม่ต้องใช้เทคนิคการเรียนรู้ของเครื่องใด ๆ อนึ่ง ในการใช้โครงข่ายประสาทเทียมที่กล่าวมา จะมีอยู่ 6 คุณลักษณะที่คณะผู้วิจัยระบุว่าจำเป็นต้องอาศัยผลลัพธ์จากคุณลักษณะตัวอื่น เช่น ความรุนแรงของการกระทำ (Result_Severity) ต้องอาศัยผลลัพธ์จาก ความสำคัญของอวัยวะ (Organ_Importance) ความรุนแรงของบาดแผล (Sore_Severity) และความรุนแรงของอาวุธ (Weapon_Severity) เฉพาะกรณีดังกล่าวนี้ จะใช้โครงข่ายประสาทเทียมแบบโมดูล (Modular neural network)

2) การทำนายคุณลักษณะทางกฎหมาย (Legal-level attribute) โดยใช้คุณลักษณะของคดีที่ได้จากขั้นตอนก่อนทั้ง 25 คุณลักษณะ ผลลัพธ์ที่ได้ คือ คุณลักษณะทางกฎหมายรวม 9 คุณลักษณะ ประกอบด้วย ระดับของเจตนาในการกระทำผิด (Act_Intention_Result) การบันดาลโทสะ (Anger) ความรู้ผิดชอบขณะกระทำผิด (Awareness) การปกปิดการกระทำ (Concealment) ความทารุณโหดร้าย (Cruelty) การไตร่ตรองไว้ก่อน (Plan) การป้องกันตัว (Prevention) ความพยายามที่เป็นไปไม่ได้ (Unachieveable Attempt) และอายุของผู้กระทำผิด (OffenderAge) ซึ่งคุณลักษณะทั้งหมดจะเชื่อมโยงกับมาตราสำคัญในประมวลกฎหมายอาญา เช่น ระดับของเจตนาในการกระทำผิด ถ้ามีระดับต่ำแสดงว่าเจตนาไม่รุนแรง เป็นเจตนาทำร้ายร่างกาย แต่ถ้ามีระดับสูงแสดงว่าร้ายแรงถึงขั้นเจตนาฆ่า หรือกรณีการไตร่ตรองไว้ก่อน ถ้าพบว่าการฆ่าผู้อื่นได้ไตร่ตรองไว้ก่อนจริง ผู้กระทำจะมีความผิดฐานฆ่าผู้อื่นโดยไตร่ตรองไว้ก่อน ฯลฯ

ทั้งนี้ มี 8 คุณลักษณะที่จะต้องเรียนรู้โดยใช้เทคนิคเอสวีเอ็ม-อาร์เอฟอี (SVM-RFE) ซึ่งเพิ่มเติมความสามารถของเอสวีเอ็มในการจัดลำดับความสำคัญของข้อมูลตัวแทน (Feature ranking) และเมื่อได้ข้อมูลตัวแทนจากอัลกอริทึมดังกล่าวแล้ว จึงนำไปเข้าสู่ต้นไม้ตัดสินใจแบบ ซี4.5 (C4.5) เพื่อให้ได้คำตอบ อนึ่ง ในผลงานวิจัยชิ้นนี้ ไม่ได้ระบุว่าคุณลักษณะเพียงหนึ่งเดียวที่ไม่ได้ผ่านเทคนิคข้างต้น คือ คุณลักษณะใด แต่ระบุเพียงว่าสามารถใช้การเขียนกฎเพื่อหาคำตอบได้เท่านั้น

หลังจากที่ได้คุณลักษณะทางกฎหมายเรียบร้อยแล้ว การระบุฐานความผิดและบทลงโทษจะใช้รูปแบบของการเชื่อมโยงคุณลักษณะทั้งหมดเข้ากับตัวบทกฎหมายที่เกี่ยวข้องผ่านฐานข้อมูล อย่างไรก็ตาม คณะผู้วิจัย ไม่ได้แสดงขั้นตอนวิธีในส่วนนี้อย่างละเอียดแต่ประการใด

สำหรับการทดลองเพื่อทดสอบประสิทธิภาพของระบบนั้น คณะผู้วิจัยได้แยกการทดลองออกเป็น 2 ส่วน ตามขั้นตอนของตัวจำแนกประเภท โดยใช้ด้วยวิธีการตรวจสอบความสมเหตุสมผลแบบข้ามส่วน 10 ส่วน แล้วหาค่าความแม่นยำเฉลี่ยจากการตรวจสอบแต่ละส่วน ทั้งนี้ ผลการทดลองพบว่า

1) การใช้พีซีเอช่วยให้การทำนายคุณลักษณะของคดีด้วยโครงข่ายประสาทเทียมโดยส่วนใหญ่ดีขึ้น โดยพบว่าโครงข่ายประสาทเทียม 11 โครงข่ายจากทั้งหมด 13 โครงข่ายที่รับข้อมูลขาเข้าซึ่งผ่านการใช้พีซีเอมาแล้วจะให้ค่าความแม่นยำสูงกว่าโครงข่ายที่รับข้อมูลขาเข้าโดยไม่ผ่านพีซีเอ นอกจากนี้ หากเป็นการทำนายคุณลักษณะที่ต้องอาศัยผลลัพธ์จากคุณลักษณะตัวอื่นแล้ว การใช้โครงข่ายประสาทเทียมแบบโมดูลช่วยให้ประสิทธิภาพสูงขึ้นกว่าการใช้โครงข่ายประสาทเทียมทำงานแยกกัน เช่น เมื่อใช้ข้อมูลขาเข้าที่ผ่านพีซีเอมาแล้ว การทำนายความรุนแรงของการกระทำ โดยใช้โครงข่ายประสาทเทียมเดิม จะได้ค่าความแม่นยำ 83.70% แต่เมื่อใช้โครงข่ายประสาทเทียมแบบโมดูล จะได้ค่าความแม่นยำสูงถึง 92.39%

2) การใช้ต้นไม้ตัดสินใจแบบ ซี4.5 และการสกัดข้อมูลตัวแทนด้วยวิธีการเอสวีเอ็ม-อาร์เอฟอี เมื่อเปรียบเทียบกับการใช้ต้นไม้ตัดสินใจแบบ ซี4.5 อย่างเดียว หรือใช้ร่วมกับการสกัดข้อมูลตัวแทนด้วยวิธีพีซีเอแล้ว ช่วยให้การทำนายคุณลักษณะทางกฎหมายดีขึ้นเพียงเล็กน้อย เพราะต้นไม้ตัดสินใจแบบ ซี4.5 ร่วมกับข้อมูลตัวแทนที่ผ่านเอสวีเอ็ม-อาร์เอฟอี สามารถชนะต้นไม้ที่นำมาเปรียบเทียบที่เหลือในการทำนายคุณลักษณะได้เพียงแค่ 3 จาก 8 คุณลักษณะเท่านั้น ส่วนมากประสิทธิภาพจะเทียบเท่ากับการใช้ต้นไม้ตัดสินใจแบบ ซี4.5 เพียงอย่างเดียว

ถึงกระนั้นก็ตาม ผลงานวิจัยค.ศ.2014 ชิ้นนี้มีข้อแตกต่างจากแบบจำลองที่ผู้วิจัยพัฒนาขึ้น กล่าวคือ

1) ชุดข้อมูลที่ใช้ในผลงานวิจัยนี้ คัดจากคำพิพากษาศาลฎีกาในคดีอาญาเพียง 150 ฉบับ ระหว่าง พ.ศ.2533 ถึงพ.ศ.2553 ในคดีความผิดเกี่ยวกับชีวิตและร่างกายตามประมวลกฎหมายอาญา ส่วนชุดข้อมูลที่ผู้วิจัยพัฒนาขึ้นเพื่อใช้กับแบบจำลอง คัดจากคำพิพากษาศาลฎีกาในคดีอาญาถึง 1,000 ฉบับ ระหว่างพ.ศ.2501 - 2559 ในคดีความผิดเกี่ยวกับชีวิตและร่างกาย ชื่อเสียง และทรัพย์

2) ผลงานวิจัยชิ้นนี้ ไม่ได้ ใช้ข้อความเป็นข้อมูลขาเข้าแต่อย่างใด โดยลักษณะข้อมูลที่ผลงานวิจัยดังกล่าวใช้นั้น ผ่านการกลั่นกรองและเข้ารหัสเอาไว้แล้วในลักษณะคล้ายกับฐานข้อมูลเอสซีดีบี ที่ได้กล่าวมาแล้วในหัวข้อที่ 3.1.2 ส่วนแบบจำลองที่ผู้วิจัยพัฒนาขึ้นจะใช้ข้อความเป็นข้อมูลขาเข้า

3) ผลงานวิจัยชิ้นนี้ ผสมผสานเทคนิคการเรียนรู้ของเครื่องเพื่อการสกัดข้อมูลตัวแทนและประมวลผลหาคำตอบ เห็นได้จากส่วนการทำนายคุณลักษณะทางกฎหมาย ซึ่งผลงานวิจัยชิ้นนี้ใช้ต้นไม้ตัดสินใจแบบ ซี4.5 ในการประมวลผล และใช้เทคนิคเอสวีเอ็ม-อาร์เอฟอีเพื่อสกัดข้อมูลตัวแทน ขณะที่แบบจำลองที่ผู้วิจัยพัฒนาขึ้นใช้เทคนิคการเรียนรู้เชิงลึกซึ่งประกอบด้วยโครงข่ายประสาทเทียมทั้งแบบวกกลับและแบบไปข้างหน้า เพื่อให้แบบจำลองสามารถสร้างข้อมูลตัวแทน และทำนายผลลัพธ์ได้ด้วยโครงข่ายเดียวกัน

Chulalongkorn University

อนึ่ง ในส่วนการสกัดคุณลักษณะของคดีที่ผลงานวิจัยชิ้นนี้ใช้โครงข่ายประสาทเทียมแบบโมดูลในการประมวลผล และใช้เทคนิคพีซีเอในการสกัดข้อมูลตัวแทนนั้น ไม่ถือว่า เป็นเทคนิคการเรียนรู้เชิงลึก เพราะไม่ได้ใช้ โครงข่ายประสาทเทียมในการสร้างข้อมูลตัวแทนตั้งแต่แรก ทว่าแบบจำลองที่ผู้วิจัยพัฒนาขึ้นได้ประยุกต์ใช้โครงข่ายประสาทเทียมแบบโมดูลตามผลงานวิจัยชิ้นนี้ด้วย

4) ผลงานวิจัยชิ้นนี้ ไม่ได้ใช้เทคนิคการเรียนรู้ของเครื่องเพื่อทำนายคำตัดสิน เพราะใช้วิธีเชื่อมโยงคุณลักษณะผ่านฐานข้อมูล ส่วนแบบจำลองที่ผู้วิจัยพัฒนาขึ้น ทำนายคำตัดสินด้วยชุดโครงข่ายประสาทเทียม

3.2.3 ผลงานวิจัยโดย พงศ์ปณัฏฐ โอสถิตย์พร, นวลวรรณ สุนทรภิษัช และ วิวัฒน์ วัฒนาวุฒิ (ค.ศ.2017) [32]

ได้ร่วมกันพัฒนาโครงสร้างการได้มาซึ่งความรู้เกี่ยวกับกฎหมายอาญาขึ้นเมื่อค.ศ.2017 โดยนำโครงสร้างความรับผิดทางอาญาตามที่เคยกล่าวมาในหัวข้อที่ 2.3 มาจัดทำเป็นออนโทโลยี (Ontology) ด้วยภาษาโอดับเบิลยูแอล (OWL ย่อมาจาก Ontology Web Language) เพื่อแสดงความสัมพันธ์ระหว่างองค์ประกอบต่าง ๆ ในโครงสร้างข้างต้น นอกจากนี้ ยังได้นำความผิดฐานฆ่าผู้อื่นประมวลกฎหมายอาญา มาตรา 288 และฐานฆ่าผู้อื่นโดยเหตุฉกรรจ์มาตรา 289 มาจัดทำเป็นกฎความสัมพันธ์บนออนโทโลยีข้างต้นด้วยภาษา เอสดับเบิลยูอาร์แอล (SWRL ย่อมาจาก Semantic Web Rule Language) เพื่อแสดงกฎการอนุมานจากออนโทโลยีที่สร้างขึ้นด้วย

อย่างไรก็ดี แม้ผลงานวิจัยชิ้นนี้จะให้ผลลัพธ์เป็นออนโทโลยีและกฎความสัมพันธ์เท่านั้น ซึ่งแตกต่างจากผลงานของผู้วิจัยซึ่งมุ่งพัฒนาแบบจำลองสำหรับทำนายผลคำตัดสินจากคำพิพากษาศาลฎีกาในคดีอาญาเป็นหลัก ก็ตาม แต่ ผลงานวิจัยชิ้นนี้และแบบจำลองที่ผู้วิจัยพัฒนาขึ้นจะใช้องค์ความรู้ตามโครงสร้างความรับผิดทางอาญาเหมือนกัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.2.4 ผลงานวิจัยโดย ธนพล ตันติศรีปรีชา และ นวลวรรณ สุนทรภิษัช (ค.ศ.2016) [33]

ได้ร่วมกันพัฒนาระบบผู้เชี่ยวชาญ (Expert system) สำหรับสำนักงานกฎหมายขึ้นเมื่อค.ศ.2016 โดยจำลองการตัดสินคดีแพ่งทั้งในเชิงเนื้อหาของข้อกฎหมายและเชิงวิธีพิจารณาความด้วยการเขียนโปรแกรมเชิงตรรกะ (Logic programming) ผ่านภาษาโปรล็อก (Prolog) ระบบนี้แบ่งการทำงานออกเป็น 2 โมดูล คือ

1) โมดูลคำอธิบาย (Explanation module) ซึ่งความจริงแล้ว คือ ส่วนต่อประสานผู้ใช้ (User Interface) เพราะเป็นโมดูลที่รับข้อมูลขาเข้าจากผู้ใช้ไปประมวลผล ก่อนนำผลลัพธ์ที่ได้มาแสดงในรูปแบบที่เป็นมิตรกับผู้ใช้ โดยระบุว่าคดีนี้ฝ่ายใดเป็นฝ่ายชนะพร้อมเหตุผลทางกฎหมายประกอบ
2) โมดูลผู้เชี่ยวชาญ (Expert module) เป็นโมดูลประมวลผลโดยรับข้อมูลขาเข้าเป็นข้อเท็จจริงที่นำเสนอโดยฝ่ายโจทก์และฝ่ายจำเลยซึ่งรับมาจากโมดูลคำอธิบาย จากนั้นนำมาผ่านกฎสำหรับค้นหาข้อเท็จจริงที่รับฟังได้เป็นที่ยุติทั้งสองฝ่าย และกฎสำหรับตัดสินคดีซึ่งอ้างอิงตามเนื้อหาในกฎหมาย

ที่เกี่ยวข้อง (เช่น หากคดีนั้นเกี่ยวข้องกับเช็ค กฎหมายที่เกี่ยวข้อง คือ ประมวลกฎหมายแพ่งและพาณิชย์) และหลักการนำเสนอพยานหลักฐาน (กฎหมายลักษณะพยาน) ตามประมวลกฎหมายวิธีพิจารณาความแพ่ง

อย่างไรก็ดี ผลงานวิจัยชิ้นนี้มุ่งเน้นการตัดสินคดีแพ่งโดยสร้างชุดของกฎทางตรรกะซึ่งแสดงผ่านการเขียนโปรแกรมเชิงตรรกะ ส่วนแบบจำลองที่ผู้วิจัยพัฒนาขึ้นจะมุ่งเน้นการตัดสินคดีอาญา โดยสร้างแบบจำลองทำนายผลคำตัดสินที่ใช้เทคนิคการเรียนรู้เชิงลึก

3.2.5 ผลงานวิจัยโดย กานกวิญจน์ โค้วสีหวัฒน์ และ พีรพล เวทีกูล (ค.ศ.2015) [8]

ได้ร่วมกันพัฒนาระบบสกัดใจความสำคัญจากคำพิพากษาศาลฎีกาขึ้นเมื่อค.ศ.2015 เพื่อเป็นระบบต้นแบบของการสกัดสารสนเทศและสรุปย่อข้อความจากคำพิพากษาศาลฎีกา ช่วยให้กระบวนการย่อข้อกฎหมายของผู้พิพากษาสะดวกรวดเร็วยิ่งขึ้นและได้ผลลัพธ์ที่ถูกต้องแม่นยำเชื่อถือได้ ตลอดจนเป็นเครื่องมือสำหรับช่วยคัดเลือกคำพิพากษาศาลฎีกาที่เป็นประโยชน์ต่อการศึกษาค้นคว้าและวิจัยทางนิติศาสตร์

ทั้งนี้ ระบบรับข้อมูลขาเข้าเป็นข้อความจากคำพิพากษาศาลฎีกา จากนั้นนำไปสกัดข้อมูลผ่านกฎที่สร้างขึ้น เพื่อกลั่นกรองข้อมูลสำคัญจากคำพิพากษาฉบับดังกล่าว ประกอบไปด้วย เลขที่คดี ประเภทคดี รายชื่อคู่ความในคดี เรื่อง องค์คณะผู้พิพากษาที่ตัดสินคดี ผู้ยื่นฎีกา และผลแห่งคำพิพากษาศาลฎีกา ตลอดจนสามารถสรุปย่อคำพิพากษาศาลฎีกาได้ออกมาเป็น 2 รูปแบบ คือ ย่อยาว ซึ่งแสดงรายละเอียดของคดีตั้งแต่ศาลชั้นต้น ศาลอุทธรณ์ และ ศาลฎีกา กับ ย่อสั้น ซึ่งแสดงเฉพาะเหตุผลของศาลฎีกาในการวินิจฉัยประเด็นแห่งคดีเท่านั้น อนึ่ง จากการทดสอบประสิทธิภาพของการสกัดข้อมูล พบว่าระบบสามารถสกัดข้อมูลโดยให้ค่าความแม่นยำมากกว่า 90% ส่วนการสรุปย่อคำพิพากษานั้น จากการวัดค่าความพึงพอใจของผู้อ่านซึ่งอยู่ในแวดวงกฎหมาย พบว่ามีค่าความพึงพอใจสูงกว่า 80%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

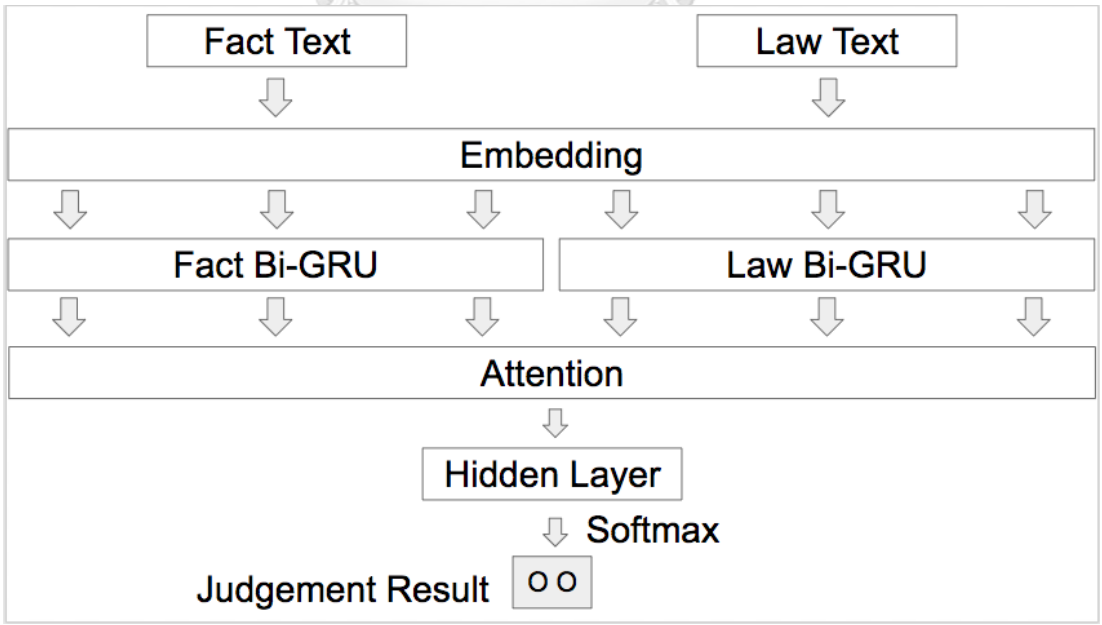
อย่างไรก็ดี ผลงานวิจัยชิ้นนี้ ไม่ได้ พัฒนาขึ้นมาเพื่อทำนายผลคำตัดสินจากคำพิพากษาแต่อย่างใด ในขณะที่แบบจำลองที่ผู้วิจัยพัฒนาขึ้นจะใช้ทำนายผลคำตัดสินจากคำพิพากษา

บทที่ 4
แนวคิดและวิธีการดำเนินงาน

4.1 แบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญาโดยใช้เทคนิคการเรียนรู้เชิงลึก

แบบจำลองนี้ มีกระบวนการทำงานเป็นไปตามรูปที่ 1.2 ในบทที่ 1 ทั้งนี้ ข้อมูลขาเข้า คือ ข้อความ 2 ชุด ชุดหนึ่ง คือ ข้อเท็จจริงจากคำพิพากษา (Fact) และอีกชุดหนึ่ง คือ ข้อกฎหมายที่ใช้กับคดีนั้น (Law) ซึ่งอ้างอิงจากฐานความผิดตามตัวบทกฎหมายโดยตรง ทั้งนี้ ข้อกฎหมายที่ใช้ **ไม่จำเป็นต้องมาจากกฎหมายที่บังคับใช้อยู่ในปัจจุบัน แม้จะเป็นกฎหมายเก่าที่ยกเลิกไปแล้วก็สามารถใช้เป็นข้อมูลขาเข้าในส่วนข้อกฎหมายได้** เพราะในบางกรณี คดีนั้นอาจต้องใช้กฎหมายเก่าบังคับ สืบเนื่องจากกฎหมายที่ออกมาใหม่ ไม่เป็นคุณแก่จำเลย เป็นการสอดคล้องกับประมวลกฎหมายอาญา มาตรา 2 วรรคหนึ่ง ที่บัญญัติว่า “บุคคลจักต้องรับโทษในทางอาญาต่อเมื่อได้กระทำการอัน**กฎหมายที่ใช้ในขณะกระทำนั้น**บัญญัติเป็นความผิดและกำหนดโทษไว้...” นั่นเอง ส่วนผลลัพธ์ของแบบจำลองจะให้ผลลัพธ์ 2 คำตอบ คือ การกระทำตามข้อเท็จจริงเป็นความผิด หรือ ไม่เป็นความผิด ตามข้อกฎหมายที่ระบุมา

สำหรับเทคนิคที่ใช้ประมวลผลในแบบจำลองนี้ คือ เทคนิคการเรียนรู้เชิงลึก โดยการต่อโครงข่ายประสาทเทียมหลายชั้น ประกอบด้วยโครงข่ายประสาทเทียมแบบวกกลับเพื่อสร้างตัวแทนของข้อความข้อเท็จจริงและข้อความข้อกฎหมาย กับโครงข่ายประสาทเทียมแบบไปข้างหน้าเพื่อทำนายผลลัพธ์การตัดสินคดี ทั้งนี้ โครงสร้างภายในแบบจำลองพิจารณาได้จากรูปที่ 4.1 ซึ่งกระบวนการภายในแบบจำลองทั้งหมดจะได้อธิบายต่อไปในบทนี้



รูปที่ 4.1 *โครงสร้างภายในแบบจำลองทำนายผลคำตัดสินคดีอาญาที่เรียนรู้จากคำพิพากษาศาลฎีกาไทย โดยใช้เทคนิคการเรียนรู้เชิงลึก*

4.2 การประมวลผลข้อมูลเบื้องต้น (Preprocessing)

ข้อมูลขาเข้าทั้งในส่วนข้อเท็จจริงและข้อกฎหมายจะผ่านการประมวลผลข้อมูลเบื้องต้น ซึ่งในที่นี้มีเพียงขั้นตอนเดียว คือ กระบวนการตัดคำ (Word segmentation) อันเป็นขั้นตอนพื้นฐานของระบบประมวลผลภาษาธรรมชาติโดยทั่วไป คำทั้งหมดที่ตัดได้จากข้อความนั้นจะนำเข้าสู่แบบจำลองโครงข่ายประสาทเทียมที่ออกแบบไว้ต่อไป

ซอฟต์แวร์ตัดคำที่ผู้วิจัยเลือกใช้มีที่มาจาก เล็กซ์โต (LexTo) [34] ซึ่งเป็นซอฟต์แวร์ที่พัฒนาขึ้นโดย ดร.ชูชาติ หฤไชยะศักดิ์ จากศูนย์วิจัยเทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (เนคเทค : NECTEC) เมื่อ พ.ศ.2549 ทั้งนี้ เล็กซ์โตใช้อัลกอริทึมตัดคำยาวที่สุด (Longest matching) โดยพิจารณาจากคำที่อยู่ในพจนานุกรมเล็กซิตรอน (Lexitron) ซึ่งเป็นพจนานุกรมเริ่มต้นของเล็กซ์โต อย่างไรก็ดี เนื่องด้วยเล็กซ์โตพัฒนาขึ้นด้วยภาษาจาวา (Java) แต่แบบจำลองที่ผู้วิจัยสร้างขึ้นพัฒนาด้วยภาษาไพธอน (Python) ดังนั้น ผู้วิจัยจึงเลือกใช้ซอฟต์แวร์ไพเล็กซ์โต (PyLexTo) [35] ซึ่งมีผู้พัฒนาขึ้นเพื่อให้สามารถใช้งานเล็กซ์โตในภาษาไพธอนได้ นอกจากนี้ เนื่องด้วยเล็กซ์โตสามารถเพิ่มเติมพจนานุกรมอื่นนอกเหนือจากพจนานุกรมเล็กซิตรอนได้เพื่อให้เล็กซ์โตรู้จักคำศัพท์ต่าง ๆ มากขึ้น ด้วยเหตุนี้ ผู้พัฒนาไพเล็กซ์โตจึงสร้างพจนานุกรมโดยรวบรวมคำศัพท์จากพจนานุกรมฉบับราชบัณฑิตยสถานและพจนานุกรมอื่น ๆ แบ่งออกเป็นหลาย ๆ ไฟล์ เพื่อนำมาใช้เป็นพจนานุกรมเพิ่มเติมด้วย

ถึงกระนั้นก็ตาม เนื่องจากแบบจำลองนี้เกี่ยวข้องกับงานด้านกฎหมาย แต่ยังไม่ปรากฏพจนานุกรมคำศัพท์กฎหมายในพจนานุกรมที่กล่าวมาข้างต้น ผู้วิจัยจึงรวบรวมคำศัพท์ทั้งหมดจากพจนานุกรมศัพท์กฎหมายไทย ฉบับราชบัณฑิตยสถาน [36] และเสริมคำศัพท์บางส่วนโดยผู้วิจัยเอง ท้ายที่สุด พจนานุกรมที่ใช้กับไพเล็กซ์โตมีจำนวนคำรวมกันทั้งหมด 88,130 คำ โดยจำนวนคำจากพจนานุกรมต่าง ๆ พิจารณาได้ตามตารางที่ 4.1

ตารางที่ 4.1 *ตารางแสดงที่มาและจำนวนของคำในพจนานุกรมที่ใช้ในซอฟต์แวร์ตัดคำ ไพเล็กซ์โต สำหรับแบบจำลองของผู้วิจัย*

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

พจนานุกรม	จำนวนคำ
เล็กซิตรอน	42,220
พจนานุกรมโดยผู้พัฒนาไพเล็กซ์โต	43,009
ศัพท์กฎหมายฉบับราชบัณฑิตยสถาน	2,769
พจนานุกรมเสริมโดยผู้วิจัยเอง	132
จำนวนคำทั้งหมด	88,130

4.3 กระบวนการทำงานภายในแบบจำลอง

แบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญาโดยใช้เทคนิคการเรียนรู้เชิงลึก สามารถแบ่งกระบวนการทำงานภายในได้ออกเป็น 4 ขั้นตอน ซึ่งแต่ละขั้นตอนอธิบายได้ตามลำดับดังนี้

4.3.1 การฝังคำ (Word embedding)

ภายหลังการตัดคำ ชุดคำของข้อความ ข้อเท็จจริง และ ข้อกฎหมาย จะนำไปผ่านการฝังคำ ซึ่งในที่นี้เลือกใช้เทคนิคตารางค้นหา [10] โดยคำแต่ละคำจะได้รับการแทนที่ หรือ “ฝัง” ด้วยรหัสตัวเลขค่าหนึ่ง รหัสข้างต้นจะสัมพันธ์กับเวกเตอร์ตัวหนึ่งที่มีค่าสุ่ม มีขนาดของเวกเตอร์ตามที่กำหนด และสามารถค้นหาได้ผ่านตารางค้นหา เวกเตอร์นี้จะกลายเป็นเวกเตอร์คำและเป็นตัวแทนของคำในการประมวลผลต่อไป

กำหนดให้ F คือ ลำดับของเวกเตอร์คำทั้งหมดในข้อเท็จจริง

L คือ ลำดับของเวกเตอร์คำทั้งหมดในข้อกฎหมาย

f_t คือ เวกเตอร์คำลำดับที่ t ของข้อเท็จจริง ซึ่งเกิดจากการฝังคำ

l_t คือ เวกเตอร์คำลำดับที่ t ของข้อกฎหมาย ซึ่งเกิดจากการฝังคำ

m คือ จำนวนคำทั้งหมดที่ตัดได้จากข้อเท็จจริง

n คือ จำนวนคำทั้งหมดที่ตัดได้จากข้อกฎหมาย

ผลลัพธ์จากการฝังคำของข้อเท็จจริงและข้อกฎหมาย แสดงได้ดังนี้

$$F = (f_1, \dots, f_m) \tag{4.1}$$

$$L = (l_1, \dots, l_n) \tag{4.2}$$

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ลำดับทั้งสองจะนำไปใช้ในขั้นตอนต่อไปเพื่อสร้างตัวแทนข้อความของข้อเท็จจริงและข้อกฎหมาย

4.3.2 การสร้างตัวแทนข้อความด้วยโครงข่ายประตูวกกลับสองทิศทาง (Bi-directional GRU : Bi-GRU)

ลำดับ F และ L เป็นข้อมูลขาเข้าในขั้นตอนนี้ โดยจะนำไปผ่านโครงข่ายประตูวกกลับสองทิศทาง หรือ Bi-GRU ซึ่งเป็นโครงข่ายประสาทเทียมแบบวกกลับที่ผู้วิจัยเลือกใช้ ทั้งนี้ จะใช้ Bi-GRU 2 โครงข่ายแยกกันสำหรับลำดับ F และ L เพื่อสร้างตัวแทนข้อความของข้อเท็จจริงและข้อกฎหมายแยกจากกัน นอกจากนี้ โครงข่ายทั้งสองจะใส่ค่าอัตราการหยุดการเรียนรู้กลางคันเอาไว้ด้วยเพื่อลดปัญหาการยึดติดข้อความที่ใช้สอนมากเกินไปขณะสอนแบบจำลองให้สร้างตัวแทนข้อความ

โดยทั่วไป โครงข่ายประสาทเทียมแบบวกกลับสองทิศทาง จะนำเวกเตอร์คำทั้งหมดของข้อความผ่านเข้าไปในโครงข่ายทั้งในลักษณะปกติและย้อนกลับ เปรียบเสมือนการอ่านข้อความจากหน้าไปหลังและย้อนจากหลังไปหน้านั่นเอง ข้อดีของการใช้โครงข่ายลักษณะนี้ คือ สามารถสร้างเวกเตอร์ที่รู้จักบริบทของคำทั้งด้านหน้าและด้านหลังได้ [24, 37] มิเช่นนั้น โครงข่ายจะรู้จักบริบทข้างหน้าเพียงอย่างเดียวเท่านั้น เทคนิคข้างต้นได้รับการนำไปใช้อย่างแพร่หลายในงานต่าง ๆ ที่ใช้เทคนิคการเรียนรู้เชิงลึก โดยไม่จำเป็นว่าต้องใช้โครงข่ายประตูวกกลับเท่านั้น [26, 38]

กำหนดให้ h_j คือ ค่าสถานะจากชั้นซ่อนของ RNN ชนิดใด ๆ สำหรับคำลำดับที่ j

$\vec{h}_j$ คือ ค่าสถานะจากชั้นซ่อนของ RNN เมื่อประมวลผลเวกเตอร์คำในข้อความจากคำแรกจนถึงคำสุดท้าย (Forward hidden state) ณ ตำแหน่งคำลำดับที่ j

$\overleftarrow{h}_j$ คือ ค่าสถานะจากชั้นซ่อนของ RNN เมื่อประมวลผลเวกเตอร์คำในข้อความจากคำสุดท้ายย้อนหาคำแรก (Backward hidden state) ณ ตำแหน่งคำลำดับที่ j

; เป็นสัญลักษณ์แสดงการเชื่อมต่อกัน (Concatenation) ระหว่างเวกเตอร์

สมการแสดงการทำงานของโครงข่ายประสาทเทียมแบบวกกลับสองทิศทาง [24] คือ

$$h_j = \left[\vec{h}_j^{\mathrm{T}}; \overleftarrow{h}_j^{\mathrm{T}}\right]^{\mathrm{T}} \tag{4.3}$$

ดังนั้น กำหนดให้

F' คือ ลำดับของค่าสถานะทั้งหมดใน Bi-GRU เมื่อป้อนเวกเตอร์คำทั้งหมดในข้อเท็จจริง

L' คือ ลำดับของค่าสถานะทั้งหมดใน Bi-GRU เมื่อป้อนเวกเตอร์คำทั้งหมดในข้อกฎหมาย

f'_t คือ ค่าสถานะจาก Bi-GRU ตามสมการ (4.3) สำหรับคำลำดับที่ t ของข้อเท็จจริง

l'_t คือ ค่าสถานะจาก Bi-GRU ตามสมการ (4.3) สำหรับคำลำดับที่ t ของข้อกฎหมาย

ผลลัพธ์จากการนำเวกเตอร์คำเข้าไปประมวลผลใน Bi-GRU แสดงได้ดังนี้

$$F' = (f'_1, \dots, f'_m) \tag{4.4}$$

$$L' = (l'_1, \dots, l'_n) \tag{4.5}$$

ทั้งนี้ เราจะถือว่า F' และ L' คือตัวแทนข้อความของข้อเท็จจริงและข้อกฎหมายตามลำดับ อย่างไรก็ดี เพื่อให้ตัวแทนข้อความมีคุณภาพมากขึ้น เราจะนำตัวแทนข้อความทั้งสองไปผ่านกลไกจุดสนใจซึ่งจะอธิบายต่อไป

4.3.3 การเพิ่มประสิทธิภาพให้ตัวแทนข้อความด้วยกลไกจุดสนใจ

การใช้กลไกจุดสนใจในขั้นตอนนี้เป็นไปเพื่อปรับปรุงตัวแทนข้อความให้สะท้อนถึงความสำคัญของคำต่าง ๆ ในข้อความนั้น เปรียบได้ดั่งการเรียนรู้ภาษาของมนุษย์ ในยามที่มนุษย์ชำนาญในการอ่านประโยคมากขึ้น มนุษย์จะทราบว่าคำใดในประโยคที่มีความสำคัญมากที่สุดต่อประโยคนั้น สำหรับข้อมูลขาเข้าในขั้นตอนนี้ คือ ลำดับ F' และ L'

ทั้งนี้ จะนำลำดับ F' และ L' มาเชื่อมต่อกันก่อน ขอเรียกว่า $[F';L']$ จากนั้นจึงคำนวณหาค่าความสนใจโดยใช้กลไกจุดสนใจในตัวเอง เพื่อหาว่าคำใดในข้อเท็จจริงและข้อกฎหมายที่มีความสำคัญสูง จากนั้นค่าความสนใจทั้งหมดที่คำนวณได้จะถูกคูณกลับเข้าไปในค่าสถานะแต่ละตัวใน $[F';L']$ ท้ายที่สุดค่าสถานะทั้งหมดที่คูณค่าน้ำหนักไว้แล้วจะบวกรวมกัน ก่อนผ่านฟังก์ชันฟังก์ชันไฮเพอร์โบลิกแทนเจนต์ เพื่อให้กลายเป็นเวกเตอร์สรุป (Summary vector) เพียงเวกเตอร์เดียว อนึ่ง กลไกจุดสนใจในตัวเองที่ใช้ในแบบจำลองนี้เลือกใช้การคำนวณแบบคูณ แสดงได้ตามฟังก์ชัน $selfAtt(x_j)$ ในสมการ (2.43)

กำหนดให้ x_j คือ ค่าสถานะลำดับที่ j จาก $[F';L']$

สมการแสดงการคำนวณหาเวกเตอร์สรุป s แสดงได้ดังนี้

$$s = \tanh\left(\sum_{j=1}^{m+n} a_j x_j\right) \quad (4.6)$$

$$a_j = softmax(selfAtt(x_j)) \quad (4.7)$$

เนื่องด้วยขั้นตอนนี้เลือกใช้กลไกจุดสนใจในตัวเองหาค่าความสำคัญของคำจากข้อเท็จจริงและข้อกฎหมายประกอบกัน จึงมีลักษณะคล้ายคลึงกับ “การปรับบทกฎหมาย” ของนักกฎหมายที่นำข้อเท็จจริงซึ่งรับฟังมาไปพิจารณาร่วมกับข้อกฎหมายที่เกี่ยวข้องเพื่อตรวจสอบว่าข้อเท็จจริงนั้นเข้าข่ายตามข้อกฎหมายดังกล่าวหรือไม่ สมมติว่ามีข้อเท็จจริงเป็นคดีฆ่าผู้อื่นโดยเจตนาตามประมวลกฎหมายอาญา มาตรา 288 การปรับบทกฎหมายจึงเป็นการตรวจสอบว่าข้อเท็จจริงดังกล่าวเข้าข่ายเป็นความผิดฐานฆ่าผู้อื่นโดยเจตนาหรือไม่นั่นเอง

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4.3.4 การทำนายผลคำตัดสินและประเด็นในคดีอาญาด้วยโครงข่ายประสาทเทียมแบบโมดูล

ขั้นตอนนี้เป็นขั้นตอนสุดท้ายของแบบจำลองโดยมีเวกเตอร์สรุป s เป็นข้อมูลขาเข้า และเทคนิคที่ใช้ คือ โครงข่ายประสาทเทียมแบบไปข้างหน้าซึ่งใช้ฟังก์ชันซอฟต์แม็กซ์เป็นฟังก์ชันกระตุ้น ผลลัพธ์ที่ได้ คือ ค่าความน่าจะเป็นของคำตอบสองคำตอบโดยที่

- ถ้าโครงข่ายนี้ใช้ทำนาย **ผลคำตัดสิน** คดีอาญา โครงข่ายจะตอบว่าการกระทำของจำเลยเป็นความผิดตามข้อกฎหมายที่ระบุมาหรือไม่ ถ้าค่าความน่าจะเป็นในฝั่งคำตอบว่า “เป็นความผิด” มีค่าสูงกว่าอีกฝั่งที่ตอบว่า “ไม่เป็นความผิด” แล้ว แสดงว่าโครงข่ายนี้ทำนายว่าการกระทำของจำเลย “เป็นความผิด”
- ถ้าโครงข่ายนี้ใช้ทำนาย **ประเด็น** ในคดีอาญาแล้ว โครงข่ายจะตอบว่าการกระทำของจำเลยเข้าหลักเกณฑ์ตามโครงสร้างความรับผิดทางอาญาในประเด็นข้อนั้นหรือไม่ เช่น โครงข่ายที่ใช้ตัดสินว่าการกระทำของจำเลย “เป็นการกระทำโดยเจตนาหรือไม่” ถ้าค่าความน่าจะเป็นในฝั่งคำตอบว่า “เจตนา” มีค่าสูงกว่าอีกฝั่งหนึ่งที่ตอบว่า “ไม่เป็นเจตนา” แล้ว แบบจำลองนี้จะทำนายว่า “จำเลยกระทำการโดยเจตนา”

สำหรับประเด็นในคดีอาญาที่แบบจำลองที่ผู้วิจัยพัฒนาขึ้นสามารถทำนายได้มีทั้งสิ้น 8 ประเด็น ประกอบด้วย

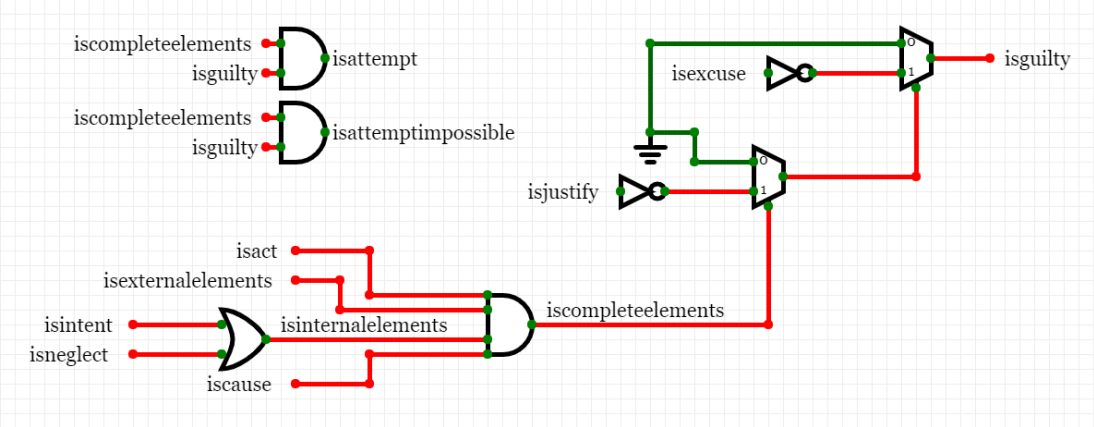
1) จำเลยมีการกระทำหรือไม่ ? หรือ $isact$
2) การกระทำของจำเลยครบองค์ประกอบภายนอกของความผิดในเรื่องนั้น ๆ แล้วหรือไม่ ? หรือ $isexternalelements$
3) การกระทำของจำเลยเป็นการกระทำโดย เจตนา หรือไม่ ? หรือ $isintent$
4) การกระทำของจำเลยเป็นการกระทำโดย ประมาท หรือไม่ ? หรือ $isneglect$
5) การกระทำของจำเลยมีเหตุยกเว้น ความผิด หรือไม่ ? หรือ $isjustify$
6) การกระทำของจำเลยมีเหตุยกเว้น โทษ หรือไม่ ? หรือ $isexcuse$
7) การกระทำของจำเลยเป็นการพยายามกระทำความผิดหรือไม่ ? หรือ $isattempt$
8) การกระทำของจำเลยเป็นการพยายามกระทำความผิดซึ่งเป็นไปไม่ได้โดยแน่แท้หรือไม่ ? หรือ $isattemptimpossible$

ทั้งนี้ การทำนายผลคำตัดสินและประเด็นในคดีอาญาข้างต้นจะมีที่มาจากลอจิกเกตแสดงโครงสร้างความรับผิดทางอาญาตามรูปที่ 2.1 ซึ่งเป็นไปตามกฎตรรกะในสมการ (2.1) (2.4) และ (2.5) อย่างไรก็ดี

1) เนื่องจากประเด็นความรับผิดโดยเด็ดขาด $isstrict$ มัก ไม่ปรากฏ ในฐานความผิดส่วนใหญ่ และไม่ได้ บันทึกไว้ในชุดข้อมูลสำหรับสอนและทดสอบแบบจำลอง จึงสามารถ ละ ประเด็นนี้ไปได้
2) เนื่องจากประเด็นการพยายามกระทำความผิด ไม่ได้ เป็นองค์ประกอบในโครงสร้างความรับผิดทางอาญา โครงข่ายที่ออกแบบไว้ในตอนแรกจึงยัง ไม่ ครอบคลุมถึง แต่เพื่อให้คำตอบของแบบจำลองในประเด็นนี้สอดคล้องกับคำตอบอื่นที่ได้มาโดยอาศัยโครงสร้างดังกล่าว ผู้วิจัยจึงดัดแปลงสมการ (2.4) และ (2.5) โดยแทนที่ประพจน์ $isfail$ และ $isalwaysfail$ ตามสมการดังกล่าวด้วยประเด็นการกระทำครบองค์ประกอบที่กฎหมายบัญญัติ $iscompleteelements$ ตามสมการ (2.2)

เพราะประเด็นนี้ผ่านการเรียนรู้จากแบบจำลองซึ่งประมวลผลข้อความข้อเท็จจริงและข้อกฎหมายที่เกี่ยวข้องในเรื่องการพยายามกระทำความผิดมาแล้ว

ดังนั้น ลอจิกเกตตามรูปที่ 2.1 จึงได้รับการดัดแปลงใหม่เป็นไปตามรูปที่ 4.2 ซึ่งโครงข่ายประสาทเทียมแบบไปข้างหน้าในแบบจำลองจะออกแบบตามวงจรที่ดัดแปลงนี้



รูปที่ 4.2 *ลอจิกเกตซึ่งได้รับการดัดแปลงเพื่อใช้ในแบบจำลอง*

ส่วนการคำนวณในประเด็นการพยายามกระทำความผิดตามสมการ (2.4) และ (2.5) ก็เปลี่ยนแปลงเป็นสมการดังต่อไปนี้ด้วย

$$isattempt = iscompleteelements \wedge isguilty \quad (4.8)$$

$$isattemptimpossible = iscompleteelements \wedge isguilty \quad (4.9)$$

ถึงกระนั้น เนื่องจากผลคำตัดสินจะอาศัยผลการทำนายประเด็นในการพิจารณาดังรูปที่ 4.2 ดังนั้น จึงออกแบบโครงข่ายประสาทเทียมแบบไปข้างหน้าในขั้นตอนนี้เป็น *โครงข่ายประสาทเทียมแบบโมดูล* ในลักษณะเดียวกันกับผลงานวิจัย [31] กล่าวคือ ผลลัพธ์ที่ได้จากโครงข่ายประสาทเทียมสำหรับทำนายประเด็นในคดีอาญา จะใช้เป็นข้อมูลขาเข้าของโครงข่ายประสาทเทียมตัวถัดไปเพื่อนำไปสู่การทำนายผลคำตัดสินในท้ายที่สุด

อนึ่ง การคำนวณฟังก์ชันการสูญเสียของโครงข่ายประสาทเทียมแบบโมดูลนี้ จะใช้วิธีรวมค่าความผิดพลาดจากการทำนายผลคำตัดสินและประเด็นทั้งหมด ก่อนดำเนินการแพร่กระจายย้อนกลับเพื่อปรับค่าน้ำหนักในโครงข่ายต่อไป ดังนั้น ลักษณะการทำงานของโครงข่ายประสาทเทียมแบบโมดูลนี้ จึงเข้าข่ายเป็น *แบบจำลองร่วม (Joint model)* กล่าวคือ เป็นโครงข่ายประสาทเทียมที่ประกอบด้วยโครงข่ายย่อย ๆ สำหรับใช้ทำนายคำตอบต่าง ๆ แต่ค่าความผิดพลาดจากการทำนายของแบบจำลองย่อยทั้งหมดจะนำมารวมคำนวณเป็นฟังก์ชันการสูญเสีย ซึ่งฟังก์ชันนี้จะนำไปใช้ปรับปรุงค่าน้ำหนักของแบบจำลองย่อยทั้งหมดร่วมกันนั่นเอง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.4 ชุดข้อมูลที่ใช้ในการทดลอง

ชุดข้อมูลที่นำมาใช้ในการทดลองนั้น ผู้วิจัยได้สร้างชุดข้อมูลทดสอบขึ้นเอง เรียกว่า ทีเอสซีซี (TSCC ย่อมาจาก Thai Supreme Court Cases) ประกอบด้วยตารางข้อมูล 2 ตาราง คือ **คำพิพากษา (Judgement)** และ **ข้อกฎหมาย (Law)** ณ ขณะที่ทำวิทยานิพนธ์นี้ ทีเอสซีซีมีข้อมูลเฉพาะคำพิพากษาในคดีอาญาเท่านั้น

การออกแบบชุดข้อมูลให้เป็นตารางข้อมูล 2 ตารางนั้นเนื่องมาจาก

1) คำพิพากษา และ ข้อกฎหมาย มีความสัมพันธ์แบบจำนวนมากต่อจำนวนมาก (Many-to-many relationship) สืบเนื่องจากคำพิพากษาฉบับหนึ่งอาจตัดสินด้วยข้อกฎหมายหลายข้อ และข้อกฎหมายข้อหนึ่งอาจใช้ตัดสินคดีได้หลายคดี ยกตัวอย่างเช่น การตัดสินคดีพยายามฆ่าผู้อื่น ต้องใช้ประมวลกฎหมายอาญา มาตรา 288 และมาตรา 80 รวม 2 ข้อกฎหมาย แต่ในทางกลับกัน คดีพยายามฆ่าผู้อื่นทุกคดีจะต้องใช้ข้อกฎหมายทั้งสองที่กล่าวมาแล้วในการตัดสินเสมอ
2) สะดวกต่อการจัดเก็บข้อมูล โดยเฉพาะข้อกฎหมายซึ่งสามารถจัดเก็บได้ทั้งข้อกฎหมายปัจจุบันและข้อกฎหมายในอดีต เพื่อให้สอดรับกับการใช้ข้อกฎหมายในขณะที่เกิดเหตุเพื่อตัดสินคดี

ข้อมูลทั้งหมดภายในชุดข้อมูลทีเอสซีซีมาจากคำพิพากษาศาลฎีกาในคดีอาญา 1,000 ฉบับที่รวบรวมจากระบบสืบค้นคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา [5] ทั้งหมดตัดสินในระหว่างปีพ.ศ.2501 - 2559 โดยเป็นคดีความผิดตามประมวลกฎหมายอาญาเพียงฉบับเดียว อย่างไรก็ดี คดีทั้งหมดจะคัดเลือกมาจากฐานความผิดเพียงบางส่วนซึ่งสามารถจำแนกได้เป็น 3 หมวดความผิด ได้แก่ ชีวิตและร่างกาย ชื่อเสียง และทรัพย์ ทั้งนี้รายละเอียดของหมวดความผิดและฐานความผิดในแต่ละหมวด แสดงตามตารางที่ 4.2

ตารางที่ 4.2 *ตารางแสดงหมวดความผิดและฐานความผิดตามประมวลกฎหมายอาญาของคำพิพากษาทั้งหมดที่นำมาใช้เป็นข้อมูลในฐานข้อมูลทีเอสซีซี*

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ลำดับที่	หมวดความผิด	ฐานความผิด
1	ชีวิตและร่างกาย	1) ฆ่าผู้อื่นโดยเจตนา (มาตรา 288) 2) ฆ่าผู้อื่นโดยมีเหตุฉกรรจ์ (มาตรา 289) เช่น ฆ่าผู้อื่นโดยไตร่ตรองไว้ก่อน 3) ทำร้ายผู้อื่นจนเป็นเหตุให้ถึงแก่ความตาย (มาตรา 290) 4) กระทำโดยประมาทจนเป็นเหตุให้ผู้อื่นถึงแก่ความตาย (มาตรา 291) 5) ทำร้ายร่างกาย (มาตรา 295) 6) ทำร้ายร่างกายโดยมีเหตุฉกรรจ์ (มาตรา 296) 7) ทำร้ายร่างกายเป็นเหตุให้ได้รับอันตรายสาหัส (มาตรา 297) 8) ทำร้ายร่างกายเป็นเหตุให้ได้รับอันตรายสาหัสโดยเหตุฉกรรจ์ (มาตรา 298)

ลำดับที่	หมวดความผิด	ฐานความผิด
		9) กระทำโดยประมาทเป็นเหตุให้ผู้อื่นได้รับอันตรายสาหัส (มาตรา 300) 10) กระทำโดยประมาทเป็นเหตุให้ผู้อื่นได้รับอันตรายแก่กายหรือจิตใจ (มาตรา 390) 11) ใช้กำลังทำร้ายผู้อื่นแต่ไม่เป็นเหตุให้ได้รับอันตรายแก่กายหรือจิตใจ (มาตรา 391)
2	ชื่อเสียง	1) หมิ่นประมาท (มาตรา 326) 2) หมิ่นประมาทด้วยการโฆษณา (มาตรา 328) 3) ดูหมิ่นผู้อื่นซึ่งหน้าหรือด้วยการโฆษณา (มาตรา 391)
3	ทรัพย์	1) ลักทรัพย์ (มาตรา 334) 2) ลักทรัพย์โดยเหตุฉกรรจ์ (มาตรา 335) เช่น ลักทรัพย์ในเวลากลางคืน 3) ลักทรัพย์ที่เป็นพระพุทธรูปหรือวัตถุในทางศาสนา (มาตรา 335 ทวิ) 4) วิ่งราวทรัพย์ (มาตรา 336) 5) กรรโชก (มาตรา 337) 6) รีดเอาทรัพย์ (มาตรา 338) 7) ชิงทรัพย์ (มาตรา 339) 8) ปล้นทรัพย์ (มาตรา 340) 9) ฉ้อโกง (มาตรา 341) 10) ฉ้อโกงโดยเหตุฉกรรจ์ (มาตรา 342) เช่น แสดงตนเป็นบุคคลอื่น 11) ฉ้อโกงประชาชน (มาตรา 343) 12) ยักยอก (มาตรา 352) 13) ผู้จัดการทรัพย์สินยักยอก (มาตรา 354) 14) ทำให้เสียทรัพย์ (มาตรา 358) 15) ทำให้เสียทรัพย์ที่ใช้หรือมีไว้เพื่อสาธารณประโยชน์ (มาตรา 360)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.4.1 การสร้างตารางข้อมูลคำพิพากษา

โดยปกติแล้ว คำพิพากษาศาลฎีกาแต่ละฉบับมักตัดสินคดีที่เกี่ยวข้องกับฐานความผิดเพียงหนึ่งฐาน แต่บางฉบับอาจตัดสินคดีในความผิดมากกว่าหนึ่งฐานก็ได้ขึ้นอยู่กับข้อเท็จจริง เช่น คดีฆ่าชิงทรัพย์ เกี่ยวข้องกับความผิดฐานฆ่าผู้อื่นและฐานชิงทรัพย์ ฯลฯ นอกจากนี้ อาจมีคดีที่เกี่ยวข้องกับความผิดฐานหนึ่ง แต่ศาลพิพากษาว่าจำเลยมีความผิดเป็นฐานอื่นด้วย เช่น โจทก์ฟ้องขอให้ลงโทษจำเลยฐานลักทรัพย์ แต่ศาลฎีกาวินิจฉัยว่าจำเลยมีความผิดฐานฉ้อโกง

จุดนี้ทำให้การวินิจฉัยว่าจำเลยมีความผิดฐานใดหรือไม่ ต้องพิจารณาแยกเป็นรายฐานความผิดไป เพราะโครงสร้างความรับผิดทางอาญาได้รับการออกแบบมาให้ตัดสินฐานความผิดได้**ครั้งละหนึ่งฐาน**เท่านั้น ด้วยเหตุนี้ การบันทึกข้อมูลคำพิพากษาแต่ละฉบับให้เป็นหนึ่งแถวข้อมูลในตารางโดยตรง จึงอาจเกิดปัญหาในการใช้งานแบบจำลองซึ่งสร้างขึ้นโดยอ้างอิงตามโครงสร้างความรับผิดทางอาญา

ด้วยเหตุนี้ ผู้วิจัยจึงสกัดคำพิพากษาที่รวบรวมมาได้ทั้งหมดให้อยู่ในรูปของ **ประเด็นปัญหา** เพื่อให้นำไปปรับใช้กับแบบจำลองได้ โดยในผลงานวิจัยชิ้นนี้ ประเด็นปัญหา หมายความว่า **ข้อเท็จจริงและข้อกฎหมายที่ใช้ตัดสินความผิดฐานหนึ่ง** โดยฐานความผิดทั้งหมดเป็นไปตามตารางที่ 4.2 ทั้งนี้ **ประเด็นปัญหาหนึ่ง ๆ ที่สกัดออกมาได้จากคำพิพากษาจะจัดเก็บเป็นแถวข้อมูลหนึ่งแถวในตาราง คำพิพากษา** นั่นเอง ทั้งนี้ คุณลักษณะทั้งหมดของแถวข้อมูลหนึ่งแถวในตารางนี้ พิจารณาได้ตามตารางที่ 4.3

ตารางที่ 4.3 *ตารางแสดงคุณลักษณะของแถวข้อมูลหนึ่งในตาราง คำพิพากษา ในชุดข้อมูลทีเอสซีซี*

ลำดับที่	ชื่อคุณลักษณะ	คำอธิบาย
1	issueid	หมายเลขไอดี (ID) ของประเด็นปัญหา เป็นคีย์หลัก (Primary key) ของแถว
2	dekaid	หมายเลขคำพิพากษาศาลฎีกา
3	year	ปีที่พิพากษาคดีในรูปแบบพุทธศักราช
4	category	หมวดความผิด ใช้อักษรแทนหมวดดังนี้ LB = ชีวิตและร่างกาย || F = ชื่อเสียง || P = ทรัพย์
5	issueno	ลำดับประเด็นปัญหาที่สกัดได้จากคำพิพากษาศาลฎีกา เริ่มต้นที่ 1 เสมอ
6	lawids	รหัสข้อกฎหมายซึ่งเป็นคีย์หลักของแถวในตาราง ข้อกฎหมาย
7	fact	ข้อความ ข้อเท็จจริง ที่สกัดได้จากย่อสั้นของคำพิพากษา
8	decision	ข้อความ คำตัดสิน ที่สกัดได้จากย่อสั้นของคำพิพากษา
9	isact	คำตอบว่า จำเลยมีการกระทำหรือไม่ ?
10	isexternalelements	คำตอบว่า การกระทำของจำเลยครบองค์ประกอบ *ภายนอก* ของความผิดตามกฎหมายที่ระบุมาหรือไม่ ?

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ลำดับที่	ชื่อคุณลักษณะ	คำอธิบาย
11	isinternalelements	คำตอบว่า การกระทำของจำเลยครบองค์ประกอบ *ภายใน* ของความผิดตามกฎหมายที่ระบุมาหรือไม่ ?
12	isintent	คำตอบว่า การกระทำของจำเลยเป็นการกระทำโดย *เจตนา* หรือไม่ ?
13	isneglect	คำตอบว่า การกระทำของจำเลยเป็นการกระทำโดย *ประมาท* หรือไม่ ?
14	iscause	คำตอบว่า การกระทำของจำเลยสัมพันธ์กับผลแห่งการกระทำนั้นหรือไม่ ?
15	isjustify	คำตอบว่า การกระทำของจำเลยมีเหตุยกเว้น *ความผิด* หรือไม่ ?
16	isexcuse	คำตอบว่า การกระทำของจำเลยมีเหตุยกเว้น *โทษ* หรือไม่ ?
17	isguilty	ผลคำตัดสินในคดีอาญาว่าจำเลยมีความผิดและต้องรับโทษหรือไม่ ?
18	isattempt	คำตอบว่า จำเลยพยายามกระทำความผิดหรือไม่ ?
19	isatttemptimpossible	คำตอบว่า จำเลยพยายามกระทำความผิดซึ่งเป็นไปไม่ได้โดยแน่แท้หรือไม่ ?

กระบวนการแปลงคำพิพากษาให้เป็นประเด็นปัญหา มีขั้นตอนดังต่อไปนี้

1) พิจารณาข้อความในย่อสั้น แล้วสกัดออกมาเป็นประเด็นปัญหาตามฐานความผิด เช่น ข้อความที่ศาลฎีกาตัดสินเกี่ยวกับความผิดฐานฆ่าผู้อื่นเพียงฐานเดียวย่อมสกัดได้ 1 ประเด็นปัญหา แต่ถ้าข้อความนั้นเกี่ยวข้องทั้งความผิดฐานฆ่าผู้อื่นและฐานชิงทรัพย์แล้ว จึงสกัดได้ 2 ประเด็นปัญหา หรือถ้าข้อความนั้นระบุว่าโจทก์ฟ้องขอให้ลงโทษฐานลักทรัพย์ แต่ศาลฎีกาตัดสินว่าเป็นฉ้อโกง ย่อมสกัดได้ 2 ประเด็นปัญหาเช่นกัน ดังนั้น จำนวนประเด็นปัญหา คือ จำนวนแถวข้อมูลที่สร้างได้จากคำพิพากษา 1 ฉบับนั่นเอง
2) บันทึก ข้อมูลเกี่ยวกับคำพิพากษา ซึ่งเป็นที่มาของประเด็นปัญหานั้น ได้แก่คุณลักษณะ issueid, dekaid, year, category และ issueno เพื่อประโยชน์ในการแบ่งข้อมูลสอนและทดสอบสำหรับแบบจำลอง และเป็นประโยชน์ต่อการอ้างอิงสำหรับนักกฎหมาย ดังนั้น ทุกแถวข้อมูลของประเด็นปัญหาที่มาจากคำพิพากษาฉบับเดียวกัน ย่อมมีข้อมูลเกี่ยวกับคำพิพากษาเหมือนกัน
3) ในแต่ละประเด็นปัญหา จะสกัดข้อความในประเด็นปัญหานั้นออกเป็น 2 คุณลักษณะ ได้แก่ **ข้อเท็จจริง (fact)** และ **คำตัดสิน (decision)** ทั้งนี้ ข้อเท็จจริง คือ ข้อความที่แสดงเหตุการณ์ต่าง ๆ อันเป็นที่มาของคดีนั้น ขณะที่คำตัดสิน คือ ข้อความที่แสดงผลคำตัดสินของศาล เช่น “จำเลยมีความผิด” หรือ “การกระทำของจำเลย ไม่เป็นความผิด” ฯลฯ อนึ่ง ในส่วนของข้อเท็จจริงนั้น หากข้อความส่วนใด ไม่ได้บ่งบอกถึงผลคำตัดสินโดยตรง แต่แสดงการใช้ดุลพินิจของศาล เช่น “ฟังได้ว่าจำเลยมีเจตนา...” หรือ “ถือไม่ได้ว่าความตายของผู้ตายเป็นผลมาจากการกระทำของจำเลย” ข้อความส่วนนี้จะครอบไว้ด้วยแท็ก <discr> ซึ่งมาจาก Discretion แปลว่า ดุลพินิจ แท็กดังกล่าวจะมีผลต่อการทดลองซึ่งจะได้กล่าวต่อไป

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4) ระบุข้อกฎหมายที่เกี่ยวข้องกับประเด็นปัญหานั้นลงในคุณลักษณะ lawids ทั้งนี้ จะระบุในรูปแบบคีย์หลักจากตาราง ข้อกฎหมาย อย่างไรก็ตาม ถ้าประเด็นปัญหานี้มีข้อกฎหมายที่เกี่ยวข้องมากกว่า 1 ข้อ จะระบุคีย์หลักทั้งหมดโดยคั่นด้วยเครื่องหมายจุลภาค (Comma) โดยใช้คีย์หลักของฐานความผิดขึ้นต้นเสมอ
5) การระบุคำตอบของผลคำตัดสินและประเด็นในคดีอาญาสำหรับประเด็นปัญหานั้น โดยพิจารณาตามโครงสร้างความรับผิดทางอาญา คำตอบเหล่านี้จะบันทึกลงในคุณลักษณะทั้งหมด *ที่ขึ้นต้นด้วยคำว่า is* โดยมีลักษณะเป็นสองคำตอบ ได้แก่ค่า 1 แทนความหมายว่า ใช่ และค่า 0 แทนความหมายว่า ไม่ใช่ อย่างไรก็ดี บางคำตอบอาจระบุเป็น -1 ถ้าประเด็นนั้น ไม่มีผลต่อการวิเคราะห์ความรับผิด เช่น หากประเด็นปัญหานั้น ระบุคำตอบของเจตนาในการกระทำความผิดว่า 1 กรณีนี้จะระบุคำตอบของความประมาทว่า -1 เพราะเจตนาและประมาทเป็นเหตุการณ์ที่ไม่เกิดร่วมกันนั่นเอง

เมื่อเสร็จสิ้นกระบวนการสกัดประเด็นปัญหาแล้ว พบว่าตาราง คำพิพากษา มีจำนวนแถวข้อมูล 1,207 แถว หรือกล่าวได้ว่ามีประเด็นปัญหา 1,207 ประเด็นนั่นเอง ทั้งนี้ สถิติที่เกี่ยวข้องกับข้อมูลในตารางนี้ แสดงอยู่ในตารางที่ 4.4 ส่วนตัวอย่างแถวข้อมูลของประเด็นปัญหาในตารางนี้ แสดงอยู่ในภาคผนวก ก.

ตารางที่ 4.4 *ตารางแสดงสถิติข้อมูลของตาราง คำพิพากษา ในชุดข้อมูลทีเอสซีซี*

หมวดความผิด	**จำนวนประเด็นปัญหา (%ของทั้งหมด)**	**จำนวนผลคำตัดสินว่าผิด (isguilty=1)**	**จำนวนผลคำตัดสินว่าไม่ผิด (isguilty=0)**	**อัตราส่วนผิด : ไม่ผิด**
ทรัพย์	473 (39%)	257	216	54 : 46
ชีวิตและร่างกาย	442 (37%)	281	161	64 : 36
ชื่อเสียง	292 (24%)	93	199	32 : 68
รวม	**1,207 (100%)**	**631**	**576**	**52 : 48**

อนึ่ง ยังมีสถิติของข้อมูลเพิ่มเติมในส่วนประเด็นในคดีอาญา ซึ่งแสดงตามตารางที่ 4.5

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.5 *ตารางแสดงสถิติข้อมูลส่วนประเด็นในคดีอาญาของตาราง คำพิพากษา ในชุดข้อมูลทีเอสซีซี*

ประเด็น	จำนวนประเด็นปัญหาที่มีคำตอบ "ใช่"	จำนวนประเด็นปัญหาที่มีคำตอบ "ไม่ใช่"	จำนวนประเด็นปัญหาที่ไม่สนใจ	อัตราส่วน ใช่ กับ ไม่ใช่
การกระทำ	1,197	10	0	99 : 1
ครบองค์ประกอบภายนอก	872	325	10	73 : 27
เจตนา	790	49	368	94 : 6
ประมาท	26	7	1,174	79 : 21
ยกเว้นความผิด	171	642	394	21 : 79
ยกเว้นโทษ	11	631	565	2 : 98
พยายาม	95	510	602	16 : 84
พยายามเป็นไปไม่ได้	13	82	1,112	14 : 86

4.4.2 การสร้างตารางข้อมูลข้อกฎหมาย

ระหว่างการสร้างแถวข้อมูลในตารางข้อมูลคำพิพากษานั้น ทุกข้อกฎหมายที่ใช้ในการตัดสินประเด็นปัญหาจะได้รับการบันทึกลงในตารางข้อกฎหมายด้วย โดยข้อกฎหมายแต่ละข้อจะมีคุณลักษณะแสดงตามตารางที่ 4.6

ตารางที่ 4.6 *ตารางแสดงคุณลักษณะของแถวข้อมูลหนึ่งในตาราง กฎหมาย ในชุดข้อมูลทีเอสซีซี*

ลำดับที่	ชื่อคุณลักษณะ	คำอธิบาย
1	lawid	หมายเลขไอดีของข้อกฎหมาย และเป็นคีย์หลักของแถว
2	lawcode	รหัสชื่อกฎหมาย ปัจจุบันมีเพียงรหัสเดียว คือ CC แทนประมวลกฎหมายอาญา (Criminal Code)
3	lawsection	มาตราของข้อกฎหมาย
4	category	หมวดหมู่ของมาตราในกฎหมายนั้น
5	effectivedate	วันที่ข้อกฎหมายนี้มีผลบังคับใช้
6	enddate	วันที่ข้อกฎหมายนี้สิ้นสุดการบังคับใช้
7	editreference	กฎหมายที่แก้ไขเพิ่มเติม หากข้อกฎหมายนั้น มาจากการแก้ไขเพิ่มเติม เช่น พ.ร.บ.แก้ไขเพิ่มเติมประมวลกฎหมายอาญา พ.ศ.2502
8	content	เนื้อความในข้อกฎหมายอ้างอิงตามตัวบทกฎหมายจริง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ทั้งนี้ ในการเก็บข้อมูลกฎหมาย จะใช้คุณลักษณะ lawid เป็นคีย์หลัก โดยตัวคีย์เกิดจากการเชื่อมข้อมูล 3 ตัวเข้าด้วยกัน ได้แก่ รหัสชื่อกฎหมาย มาตราของกฎหมาย และจำนวนครั้งในการแก้ไข เช่น

- “CC-288-00” หมายความว่า ประมวลกฎหมายอาญา มาตรา 288 ซึ่งบัญญัติขึ้นเป็นครั้งแรก
- “CC-335-02” หมายความว่า ประมวลกฎหมายอาญา มาตรา 335 แก้ไขเพิ่มเติมครั้งที่ 2

ข้อดีของการกำหนดรูปแบบ lawid ข้างต้น คือ

1) สื่อความหมายให้เข้าใจได้ง่าย เป็นประโยชน์ต่อการอ้างอิงสำหรับนักกฎหมาย
2) ทำให้ตาราง ข้อกฎหมาย สามารถจัดเก็บตัวบทกฎหมายได้ทั้งตัวบทเดิมและตัวบทปัจจุบัน เพื่อให้สอดคล้องกับแนวคิดของแบบจำลองที่สามารถใส่ข้อกฎหมายที่ใช้บังคับขณะที่คดีเกิดขึ้นได้ แม้จะเป็นข้อกฎหมายเก่าที่ไม่ได้ใช้บังคับในปัจจุบันแล้วก็ตาม

ภายหลังการกำหนด lawid แล้ว จะบันทึกข้อความที่ปรากฏในตัวบทกฎหมายนั้นลงในคุณลักษณะ content ก่อนบันทึกข้อมูลอื่นที่เกี่ยวข้อง ได้แก่ หมวดหมู่ วันที่มีผลใช้บังคับ วันสิ้นสุดการใช้บังคับ และกฎหมายที่แก้ไขเพิ่มเติม เป็นอันเสร็จสิ้นการบันทึกข้อมูลหนึ่งแถว

เมื่อกระบวนการรวบรวมและสร้างข้อมูลในตาราง ข้อกฎหมาย เสร็จสิ้น ตารางนี้ประกอบด้วยฐานความผิดตามภาค 2 ความผิด และภาค 3 ลหุโทษ แห่งประมวลกฎหมายอาญา รวมถึงบทบัญญัติทั่วไปในภาค 1 ซึ่งอาจต้องนำมาใช้ร่วมด้วยโดยผลของการวิเคราะห์คดีอาญาด้วยโครงสร้างความรับผิดทางอาญา สำหรับสถิติที่เกี่ยวข้องกับข้อมูลในตารางข้อกฎหมาย แสดงตามตารางที่ 4.7 สำหรับตัวอย่างแถวข้อมูลของตารางนี้ แสดงอยู่ในภาคผนวก ก.

ตารางที่ 4.7 *ตารางแสดงสถิติข้อมูลของตาราง ข้อกฎหมาย ในชุดข้อมูลทีเอสซีซี*

หมวดหมู่ในประมวลกฎหมายอาญา	จำนวนมาตราในกฎหมาย	จำนวนข้อกฎหมายที่สกัดได้
ภาค 1 บทบัญญัติทั่วไป	19	28
ภาค 2 ความผิด และภาค 3 ลหุโทษ	43	94
รวม	**62**	**122**

บทที่ 5
การทดลองและผลการทดลอง

5.1 สมมติฐานการทดลอง

การทดลองนี้มุ่งวัดประสิทธิภาพของแบบจำลองที่พัฒนาขึ้นด้วยเทคนิคการเรียนรู้เชิงลึกตามบทที่ 4 ซึ่งต่อไปจะขอเรียกว่า **แบบจำลองที่นำเสนอ** ทั้งนี้ การทดลองจะทดสอบสมมติฐานดังต่อไปนี้

1) แบบจำลองที่นำเสนอมีประสิทธิภาพในการทำนายผลคำตัดสินและประเด็นในคดีอาญาสูงกว่าการใช้เทคนิคการเรียนรู้ของเครื่องเดิม เมื่อวัดจากค่าความแม่นยำและค่าเฉลี่ยมหภาคของเอฟวัน
2) การใช้เวกเตอร์ของโครงข่ายประตูวกกลับสองทิศทางเป็นเวกเตอร์ตัวแทนข้อความ ทำให้ประสิทธิภาพการทำนายผลคำตัดสินของแบบจำลองดีกว่าการใช้ตัวแทนข้อความดั้งเดิมที่นิยมใช้ในงานประมวลผลภาษาธรรมชาติ ทั้งนี้ เลือกตัวแทนข้อความดั้งเดิมเป็นแบบจำลองทีเอฟไอดีเอฟ
3) การใช้กลไกจุดสนใจช่วยเพิ่มประสิทธิภาพให้เวกเตอร์ของโครงข่ายประตูวกกลับสองทิศทาง ทำให้ประสิทธิภาพการทำนายผลคำตัดสินของแบบจำลองดีกว่าการใช้เวกเตอร์ของโครงข่ายอย่างเดียว
4) การใช้กลไกจุดสนใจกับเวกเตอร์ข้อเท็จจริงและข้อกฎหมายร่วมกันตามแบบจำลองที่นำเสนอให้ประสิทธิภาพการทำนายผลคำตัดสินของแบบจำลองดีกว่าการใช้กลไกจุดสนใจในตัวเองแบบแยกข้อความ
5) การใช้โครงข่ายประสาทเทียมแบบโมดูลในการทำนายผลคำตัดสินสามารถลอกเลียนโครงสร้างความรับผิดทางอาญาได้ โดยมีประสิทธิภาพดีกว่าการใช้กฎทางตรรกะประมวลผล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.2 การออกแบบการทดลอง

5.2.1 การแบ่งชุดข้อมูล

การทดลองนี้จัดทำบนชุดข้อมูลทีเอสซีซี โดยแบ่งชุดข้อมูลออกเป็น 3 ส่วน คือ ชุดข้อมูลสอน (Training set) ชุดข้อมูลพัฒนา (Validation / Development set) และชุดข้อมูลทดสอบ (Test set) โดยจะแบ่งให้ได้อัตราส่วน 60 : 20 : 20 สำหรับชุดข้อมูลสอน ชุดข้อมูลพัฒนา และชุดข้อมูลทดสอบตามลำดับ

อย่างไรก็ดี การแบ่งชุดข้อมูลด้วยวิธีสุ่ม ไม่สามารถ ทำได้ มิเช่นนั้น อาจเกิดเหตุการณ์ที่แบบจำลองซึ่งเรียนรู้จากคำพิพากษาใหม่ถูกนำไปใช้ทำนายคำพิพากษาในอดีตที่อยู่ในชุดข้อมูลทวนสอบหรือชุดข้อมูลทดสอบ ซึ่งขัดตรรกะของนักกฎหมายที่ใช้คำพิพากษาในอดีตเพื่อวิเคราะห์คดีที่กำลังเกิดอยู่ในปัจจุบัน ดังนั้น การแบ่งข้อมูลจะแยกตามปีที่ตัดสินคดี เพื่อรับประกันว่า คำพิพากษาที่ใช้สอนแบบจำลอง คือ คำพิพากษาในอดีตทั้งหมด และคำพิพากษาที่ใช้ทดสอบประสิทธิภาพ คือ คำพิพากษาที่ใหม่กว่า ทั้งนี้ ผู้วิจัยได้กำหนดปีที่ใช้ตัดแบ่งชุดข้อมูลโดยพยายามรักษาอัตราส่วนของชุดข้อมูลที่กล่าวมาให้ได้ใกล้เคียงที่สุด

สุดท้ายแล้ว สามารถตัดแบ่งชุดข้อมูลได้ โดยมีรายละเอียดปรากฏตามตารางที่ 5.1 อนึ่ง ชุดข้อมูลทั้งหมดไม่ผ่าน การเติมข้อมูลสมมติ (Upsampling / Downsampling) เพื่ออัตราส่วนคำตอบที่เท่ากัน

ตารางที่ 5.1 *ตารางแสดงสถิติชุดข้อมูลที่แบ่งออกจากชุดข้อมูลทีเอลซีซีเพื่อการทดลองทดสอบประสิทธิภาพของแบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญา*

ชุดข้อมูล (ช่วงพ.ศ. ที่แบ่ง)	**จำนวนประเด็นปัญหา (%ชุดข้อมูล)**	**จำนวนผลคำตัดสินว่าผิด (isguilty=1)**	**จำนวนผลคำตัดสินว่าไม่ผิด (isguilty=0)**	**อัตราส่วนผิด : ไม่ผิด**
สอน (2501 - 2538)	694 (58%)	324	370	47 : 53
พัฒนา (2539 - 2553)	267 (22%)	158	109	59 : 41
ทดสอบ (2554 – 2559)	246 (20%)	149	97	61 : 39
รวม	**1,207 (100%)**	**631**	**576**	**52 : 48**

5.2.2 แบบจำลองที่นำเสนอซึ่งใช้ในการทดลอง

แบบจำลองที่นำเสนอนี้พัฒนาขึ้นด้วยภาษาไพธอน โดยใช้ไลบรารี่ไพทอร์ช (PyTorch) ในการสร้างโครงข่ายประสาทเทียมทั้งหมด ส่วนกฎการสอนนั้น ใช้ค่าติดลบลอการิทึมของความเป็นไปได้เป็นฟังก์ชันการสูญเสีย โดยคำนวณจากความผิดพลาดรวมในการทำนายผลคำตัดสินและทุกประเด็นในคดีอาญา ส่วนวิธีการหาค่าเหมาะที่สุดใช้เอสจีดี

CHULALONGKORN UNIVERSITY

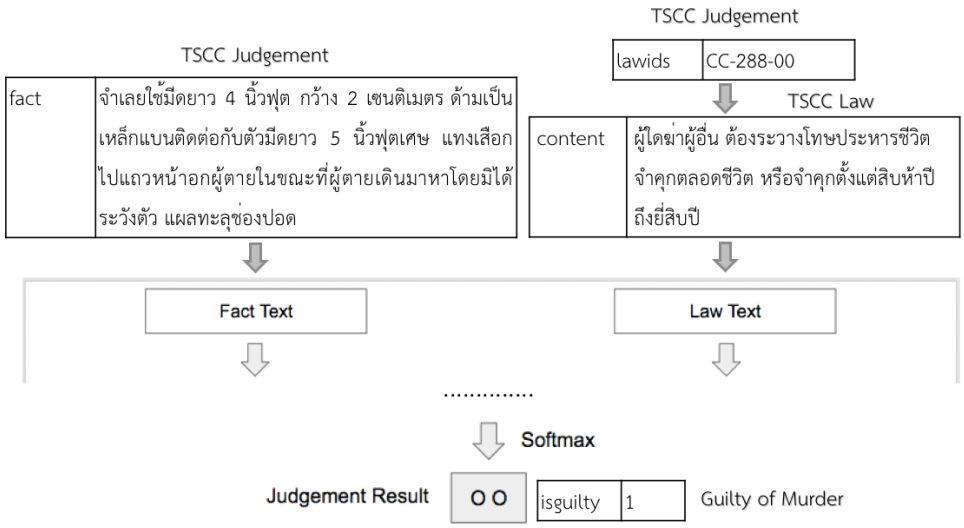
สำหรับการสอนข้อมูลให้แบบจำลองซึ่งรับข้อมูลขาเข้าเป็นข้อเท็จจริง และข้อกฎหมายนั้น

- ข้อเท็จจริง มาจากคุณลักษณะ content ของตาราง คำพิพากษา ถ้าข้อความในคุณลักษณะ content มีแท็ก <discr> อยู่ด้วยแล้ว จะตัดข้อความในแท็กดังกล่าวออกไป
- ข้อกฎหมาย มาจากคุณลักษณะ lawids ของตารางคำพิพากษา ซึ่งถอดรหัสเป็นข้อความข้อกฎหมายด้วยการเชื่อม (Map) เข้าหาคุณลักษณะ lawid ของตาราง ข้อกฎหมาย เพื่อเข้าไปดึงข้อความจากคุณลักษณะ content ของตารางดังกล่าวต่อไป อนึ่ง ถ้าคุณลักษณะ lawids ประกอบด้วยข้อกฎหมายหลายข้อที่เชื่อมกันด้วยเครื่องหมายจุลภาค เช่น “CC-288-00,CC-080-00” ข้อความกฎหมายแต่ละข้อจะต่อเชื่อมกันด้วยช่องว่าง (Whitespace) ก่อนนำไปใช้เป็นข้อกฎหมายเพื่อนำเข้าแบบจำลองต่อไป

สำหรับข้อมูลคำตอบที่จะสอนให้แบบจำลองนี้

- ผลคำตัดสินมาจากคุณลักษณะ isguilty ของตาราง คำพิพากษา
- ประเด็นในคดีอาญามาจากคุณลักษณะทั้งหมด*ที่ขึ้นต้นด้วยคำว่า is* ของตาราง คำพิพากษา

กระบวนการข้างต้นสามารถพิจารณาเป็นภาพได้ตามรูปที่ 5.1



รูปที่ 5.1 *วิธีการใช้งานแบบจำลองทำนายผลคำตัดสินคดีอาญาที่เรียนรู้จากคำพิพากษาศาลฎีกาไทย โดยใช้เทคนิคการเรียนรู้เชิงลึก ร่วมกับชุดข้อมูล ทีเอสซีซี ทั้งนี้ แสดงเฉพาะส่วนการทำนายผลคำตัดสินเท่านั้น*

เมื่อสอนแบบจำลองด้วยชุดข้อมูลสอนเรียบร้อยแล้ว เราจะทดสอบแบบจำลองนั้นกับชุดข้อมูลพัฒนา และปรับค่าพารามิเตอร์ที่เกี่ยวข้องกับโครงข่าย (Hyperparameter Tuning) จนกว่าจะได้แบบจำลองที่ให้ค่าความแม่นยำสูงที่สุด แบบจำลองที่ให้ค่าความแม่นยำสูงสุดจะถือเป็นแบบจำลองที่สมบูรณ์ซึ่งจะนำไปใช้ทดสอบกับชุดข้อมูลทดสอบต่อไป อย่างไรก็ตาม มีพารามิเตอร์บางค่าที่กำหนดให้เป็น **ค่าคงที่** ดังนี้

- จำนวนข้อมูลย่อย (Batch Size) เพื่อสอนข้อมูล มีค่าเป็น **10**
- อัตราการหยุดการเรียนรู้กลางคันของโครงข่ายประตูวกกลับสองทิศทาง เท่ากับ **0.2**
- ขนาดของเวกเตอร์ผลลัพธ์จากโครงข่ายประตูวกกลับสองทิศทาง เท่ากับ ขนาดของโครงข่ายประสาทเทียมแบบไปข้างหน้าที่ใช้ทำนายผลคำตัดสินและประเด็น

สำหรับชุดพารามิเตอร์ที่ทดสอบ และค่าที่ดีที่สุดของแต่ละพารามิเตอร์ พิจารณาได้ตามตารางที่ 5.2

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 5.2 *ตารางแสดงพารามิเตอร์ที่ทดลองปรับกับแบบจำลอง และชุดพารามิเตอร์ที่ทำให้แบบจำลองได้ค่าความแม่นยำสูงที่สุดในชุดข้อมูลพัฒนา*

พารามิเตอร์	ค่าที่ทดลอง	ค่าที่ดีที่สุด
อัตราการเรียนรู้	0.005, 0.01	0.01
ขนาดของเวกเตอร์คำ	50, 100	100
จำนวนชั้นของโครงข่ายประตูวกกลับสองทิศทาง	1, 2	2
ขนาดของชั้นซ่อนในโครงข่ายประสาทเทียมแบบไปข้างหน้าที่ใช้ทำนายผลคำตัดสินและประเด็น	50, 100	50

5.2.3 แบบจำลองตั้งต้น (Baseline model)

แบบจำลองตั้งต้นที่นำมาใช้ประกอบการทดลองนี้ แบ่งได้เป็น 3 กลุ่ม คือ แบบจำลองที่ใช้เทคนิคการเรียนรู้ของเครื่องเดิม แบบจำลองที่ใช้เทคนิคการเรียนรู้เชิงลึก และแบบจำลองที่ใช้เทคนิคการเรียนรู้เชิงลึกร่วมกับกฎตรรกะ

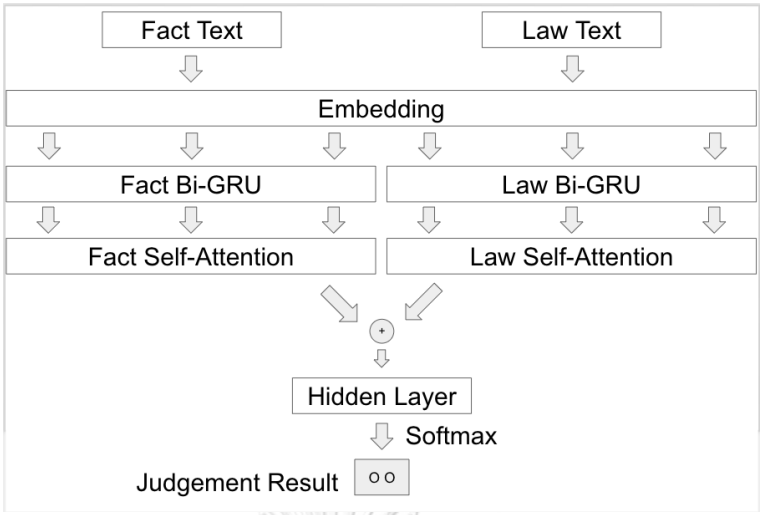
แบบจำลองที่ใช้เทคนิคการเรียนรู้ของเครื่องเดิม ใช้เพื่อตรวจสอบสมมติฐานข้อ 1 ประกอบด้วย

1) เนอีฟเบยส์ (Naïve Bayes) แบบจำลองหนึ่งที่นิยมในงานจำแนกประเภทข้อความ (Text Classification)
2) เอสวีเอ็ม ซึ่งใช้งานในผลงานวิจัย [27, 31] อย่างไรก็ดี เอสวีเอ็มต้องผ่านการปรับค่าพารามิเตอร์ที่เกี่ยวข้องด้วยเหมือนแบบจำลองที่นำเสนอ ทั้งนี้ จะใช้เอสวีเอ็มที่มีค่าพารามิเตอร์ C เป็น 0.25 เพราะให้ค่าความแม่นยำสูงสุดกับชุดข้อมูลพัฒนา

แบบจำลองที่ใช้เทคนิคการเรียนรู้เชิงลึก ประกอบด้วย

1) แบบจำลองที่นำเสนอซึ่งตัดกลไกจุดสนใจออก ขอเรียกว่า **แบบจำลองไม่มีกลไกจุดสนใจ** ซึ่งเป็นผลให้ภายหลังได้เวกเตอร์ข้อความจากโครงข่ายประตูวกกลับสองทิศทางแล้ว เวกเตอร์ข้อเท็จจริงและข้อกฎหมายจะนำไปต่อเชื่อมกันเป็นเวกเตอร์สรุป แล้วนำไปใช้ทำนายผลคำตัดสินและประเด็นในคดีอาญาทันที เราใช้แบบจำลองนี้เพื่อตรวจสอบสมมติฐานข้อ 2 และ 3
2) แบบจำลองที่นำเสนอซึ่งใช้กลไกจุดสนใจในตัวเองแบบแยกข้อความกัน โดยขอเรียกว่า **แบบจำลองจุดสนใจในตัวเองแยกข้อความ** กล่าวคือ ภายหลังได้เวกเตอร์ข้อความจากโครงข่ายประตูวกกลับสองทิศทางแล้ว เวกเตอร์ข้อเท็จจริงและข้อกฎหมายจะนำไปผ่านกลไกจุดสนใจในตัวเองแบบแยกต่างหากจากกัน แล้วจึงนำมาต่อเชื่อมเพื่อใช้เป็นเวกเตอร์สรุปต่อไป เราใช้แบบจำลองนี้ตรวจสอบสมมติฐานข้อ 3 และข้อ 4 อนึ่ง กระบวนการทำงานของแบบจำลองนี้ แสดงตามรูปที่ 5.2

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 5.2 *โครงสร้างภายในแบบจำลองจุดสนใจในตัวเองแยกข้อความ*

อนึ่ง แบบจำลองทั้งสองจะผ่านการปรับพารามิเตอร์เพื่อหาค่าที่ดีที่สุดสำหรับแบบจำลองนั้น โดยเลือกใช้ค่าคงที่ชุดเดียวกับแบบจำลองที่นำเสนอ ทั้งนี้ ชุดพารามิเตอร์ที่ดีที่สุดของแบบจำลองทั้งสองแสดงอยู่ในตารางที่ 5.3 อย่างไรก็ดี สำหรับแบบจำลองที่ใช้กฎตรรกะนั้นจะยึดค่าพารามิเตอร์ตามแบบจำลองเดิม

ตารางที่ 5.3 *ตารางแสดงพารามิเตอร์ที่ดีที่สุดสำหรับแบบจำลองไม่มีกลไกจุดสนใจและแบบจำลองจุดสนใจในตัวเองแยกข้อความ ซึ่งใช้เป็นแบบจำลองตั้งต้นในการทดลอง*

พารามิเตอร์	แบบจำลองไม่มีกลไกจุดสนใจ	แบบจำลองจุดสนใจในตัวเองแยกข้อความ
อัตราการเรียนรู้	0.01	0.01
ขนาดของเวกเตอร์คำ	100	50
จำนวนชั้นของโครงข่ายประตูวกกลับสองทิศทาง	1	1
ขนาดของชั้นซ่อนในโครงข่ายประสาทเทียมแบบไปข้างหน้าที่ใช้ทำนายผลคำตัดสินและประเด็น	100	50

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

แบบจำลองที่ใช้เทคนิคการเรียนรู้เชิงลึกร่วมกับกฎตรรกะ คือ การนำแบบจำลองตั้งต้นที่ใช้เทคนิคการเรียนรู้เชิงลึกซึ่งได้เอ่ยไปแล้วรวมถึงแบบจำลองที่นำเสนอมาดัดแปลง โดยเมื่อได้ผลการทำนายประเด็นในคดีอาญาจากโครงข่ายประสาทเทียมแบบไปข้างหน้าแล้ว จะนำไปผ่านลอจิกเกตตามรูปที่ 4.2 ซึ่งแสดงได้เป็นกฎตรรกะตามสมการ (2.1) (4.8) และ (4.9) เพื่อทำนายผลคำตัดสินแทนการใช้โครงข่ายประสาทเทียมแบบโมดูล ทั้งนี้ เพื่อทดสอบสมมติฐานข้อ 5

5.3 ผลการทดลอง

สำหรับหัวข้อนี้ จะขอใช้ตัวแปร Acc แทนค่าความแม่นยำ (ค่าเฉลี่ยจุลภาคของเอฟวัน) และ $F1$ แทนค่าเฉลี่ยมหภาคของเอฟวัน

5.3.1 การทำนายผลคำตัดสิน

ผลการทดลองส่วนนี้เป็นไปตามตารางที่ 5.4 โดยพบว่า

- ในภาพรวม แบบจำลองที่นำเสนอได้ Acc และ $F1$ สูงที่สุดในบรรดาแบบจำลองอื่นทั้งหมด โดยอยู่ที่ 64.63% และ 63.82% ตามลำดับ
- ในการทำนายคำตัดสินว่า “ผิด” แบบจำลองเอสวีเอ็มได้ Acc และ $F1$ สูงที่สุด โดยอยู่ที่ 78.52% และ 72.67% ตามลำดับ
- ในการทำนายคำตัดสินว่า “ไม่ผิด” แบบจำลองกลไกจุดสนใจในตัวเองแยกข้อความได้ Acc และ $F1$ สูงที่สุด โดยอยู่ที่ 78.35% และ 62.81% ตามลำดับ

ตารางที่ 5.4 *ตารางแสดงผลการทดลองวัดประสิทธิภาพการทำนายผลคำตัดสินในคดีอาญาจากแบบจำลองต่าง ๆ ผลที่ดีที่สุดจะแสดงเป็น**<u>ตัวหนาและขีดเส้นใต้</u>***

แบบจำลอง	ทำนายว่า ”ผิด“		ทำนายว่า ”ไม่ผิด“		ภาพรวม	
	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)
เทคนิคการเรียนรู้ของเครื่องเดิม						
เนอีฟเบยส์	49.66	58.50	69.07	56.07	57.32	57.28
เอสวีเอ็ม	**<u>78.52</u>**	**<u>72.67</u>**	42.27	48.24	64.23	60.45
เทคนิคการเรียนรู้เชิงลึก						
ไม่มีกลไกจุดสนใจ	65.10	68.31	60.82	56.73	63.41	62.52
จุดสนใจในตัวเองแยกข้อความ	53.69	64.00	**<u>78.35</u>**	**<u>62.81</u>**	63.41	63.40
แบบจำลองที่นำเสนอ	65.77	69.26	62.89	58.37	**<u>64.63</u>**	**<u>63.82</u>**
เทคนิคการเรียนรู้เชิงลึก ร่วมกับกฎตรรกะ						
ไม่มีกลไกจุดสนใจ	48.99	57.03	64.95	53.39	55.28	55.21
จุดสนใจในตัวเองแยกข้อความ	63.76	63.33	42.27	42.71	55.28	53.02
แบบจำลองที่นำเสนอ	48.32	56.69	65.98	53.78	55.28	55.24

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ข้อพิจารณาประการแรกจากผลการทดลอง คือ แบบจำลองที่นำเสนอซึ่งใช้เทคนิคการเรียนรู้เชิงลึกสามารถทำนายผลลัพธ์ได้ดีกว่าแบบจำลองที่ใช้เทคนิคการเรียนรู้ของเครื่องเดิม เพราะมีค่า Acc และ $F1$ ในภาพรวมสูงกว่าเนอีฟเบยส์และเอสวีเอ็ม นอกจากนี้ แบบจำลองที่นำเสนอรวมถึงแบบจำลองตั้งต้นที่ใช้เทคนิคการเรียนรู้เชิงลึกซึ่งไม่ใช้กฎตรรกะ ยังทำนายผลลัพธ์ได้ดีกว่าแบบจำลองที่ใช้เทคนิคการเรียนรู้ของเครื่องเดิมด้วย เมื่อพิจารณาจากค่า $F1$

อนึ่ง แม้พิจารณาประสิทธิภาพแยกประเภทคำตอบแล้วพบว่า แบบจำลองที่นำเสนอยัง ไม่มี ประสิทธิภาพสูงที่สุดในบรรดาแบบจำลองตั้งต้นที่ใช้ทดสอบก็ตาม ทว่าแบบจำลองที่ให้ประสิทธิภาพสูงสุดเมื่อทำนายผลคำตัดสินแยกประเภทนั้นเกิดความโน้มเอียงในการทำนายผลคำตัดสิน พิจารณาได้จากจำนวนคำตอบที่แบบจำลองแต่ละแบบทำนายได้เปรียบเทียบกับจำนวนคำตอบที่แบบจำลองตอบได้ถูกต้องตามตารางที่ 5.5 ซึ่งพบว่า

1) แบบจำลองที่มีประสิทธิภาพสูงสุดในการทำนายผลคำตัดสินว่า “ผิด” ได้แก่ เอสวีเอ็ม ทำนายผลคำตัดสินว่า “ผิด” ถึง 173 ประเด็นปัญหาจากทั้งหมด 246 ประเด็นปัญหา คิดเป็น 70% ของประเด็นปัญหาทั้งหมด นอกจากนี้ แม้ค่า Acc และ $F1$ เมื่อทำนายว่า “ผิด” ของเอสวีเอ็มจะสูงที่สุด แต่ค่า Acc และ $F1$ ของเอสวีเอ็มเมื่อทำนายว่า “ไม่ผิด” กลับต่ำที่สุดในบรรดาแบบจำลองทั้งหมด
2) แบบจำลองที่มีประสิทธิภาพสูงสุดในการทำนายผลคำตัดสินว่า “ไม่ผิด” ได้แก่ แบบจำลองจุดสนใจในตัวเองแยกข้อความ แม้จะทำนายผลคำตัดสินว่า “ไม่ผิด” 145 ประเด็นปัญหาจากทั้งหมด 246 ประเด็นปัญหา คิดเป็น 59% ของประเด็นปัญหาทั้งหมดก็ตาม แต่แบบจำลองนี้มีคำเฉลยที่ทายว่า “ไม่ผิด” เพียง 97 ประเด็นปัญหา คิดเป็น 39% ของประเด็นปัญหาเท่านั้น แบบจำลองนี้จึงทำนายเกินกว่าสัดส่วนเฉลยที่แท้จริงถึง 20%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ข้อพิจารณาประการที่สอง คือ การใช้โครงข่ายประตูวกกลับสองทิศทางเป็นตัวแทนข้อความในแบบจำลองที่ใช้เทคนิคการเรียนรู้เชิงลึกช่วยให้ประสิทธิภาพของแบบจำลองที่ใช้เทคนิคการเรียนรู้เชิงลึกสูงกว่าแบบจำลองที่ใช้เทคนิคการเรียนรู้ของเครื่องเดิม เพราะหากเปรียบเทียบแบบจำลองไม่มีกลไกจุดสนใจกับเนอีฟเบยส์และเอสวีเอ็ม จะพบว่าแบบจำลองไม่มีกลไกจุดสนใจ มีค่า $F1$ สูงกว่า นอกจากนี้ การใช้ตัวแทนข้อความดังกล่าวยังช่วยป้องกันปัญหาตัวแทนข้อความอ่อนไหว (Sensitive) ต่อแบบจำลองด้วย ดังจะเห็นได้จากผลต่างค่า Acc ระหว่างการทำนายผลคำตัดสินว่า “ผิด” และ “ไม่ผิด” ในเอสวีเอ็มซึ่งสูงถึง 36.25% และในส่วนของเนอีฟเบยส์สูงถึง 19.41% ทว่าผลต่างค่า Acc ของแบบจำลองไม่มีกลไกจุดสนใจ มีค่าเพียง 4.28% เท่านั้น

ตารางที่ 5.5 *ตารางจำนวนคำตอบที่แบบจำลองแต่ละแบบทำนายได้เปรียบเทียบกับจำนวนคำตอบที่แบบจำลองตอบได้ถูกต้อง ผลที่ดีที่สุดจะแสดงเป็น**ตัวหนาและขีดเส้นใต้***

แบบจำลอง	ทำนายว่า "ผิด"			ทำนายว่า "ไม่ผิด"		
	เฉลย	ทำนาย	ถูกต้อง	เฉลย	ทำนาย	ถูกต้อง
เทคนิคการเรียนรู้ของเครื่องเดิม						
เนอีฟเบยส์	149	104	74	97	142	67
เอสวีเอ็ม		173	**117**		73	41
เทคนิคการเรียนรู้เชิงลึก						
ไม่มีกลไกจุดสนใจ	149	135	97	97	111	59
จุดสนใจในตัวเองแยกข้อความ		101	80		145	**76**
แบบจำลองที่นำเสนอ		134	98		112	61
เทคนิคการเรียนรู้เชิงลึก ร่วมกับกฎตรรกะ						
ไม่มีกลไกจุดสนใจ	149	107	73	97	139	63
จุดสนใจในตัวเองแยกข้อความ		151	95		95	41
แบบจำลองที่นำเสนอ		105	72		141	64

อย่างไรก็ดี แม้แบบจำลองที่นำเสนอจะมีผลต่างค่า Acc เพียง 2.88% ซึ่งน้อยที่สุดในบรรดาแบบจำลองทั้งหมดก็ตาม แต่แบบจำลองจุดสนใจในตัวเองแยกข้อความกลับมีผลต่างค่า Acc สูงถึง 24.66% มากกว่าแบบจำลองไม่มีกลไกจุดสนใจและเนอีฟเบยส์ แสดงว่าแม้การใช้โครงข่ายประตูวกกลับสองทิศทางเป็นตัวแทนข้อความจะ ไม่ ทำให้เกิดความอ่อนไหวของตัวแทนข้อความเลยก็ตาม แต่หากมีเวกเตอร์ข้อความมากกว่า 1 เวกเตอร์ซึ่งเติมกลไกจุดสนใจแยกต่างหากจากกันแล้วนำไปต่อเชื่อมกันแล้ว จะทำให้เกิดปัญหาตัวแทนข้อความอ่อนไหวต่อแบบจำลองได้

ข้อพิจารณาประการที่สาม คือ การใช้กลไกจุดสนใจช่วยให้ประสิทธิภาพของแบบจำลองดีขึ้น เห็นได้จากแบบจำลองที่ใช้กลไกจุดสนใจทั้งสองแบบมีค่า $F1$ สูงกว่าแบบจำลองไม่มีกลไกจุดสนใจ ทั้งนี้ น่าสังเกตว่า ค่า Acc ของแบบจำลองที่นำเสนอ เมื่อทำนายผลคำตัดสินว่า "ผิด" มีค่าสูงที่สุดในแบบจำลองที่ใช้เทคนิคการเรียนรู้เชิงลึกทั้งสาม แต่ในการทำนายผลคำตัดสินว่า "ไม่ผิด" นั้น ค่า Acc ของแบบจำลองจุดสนใจในตัวเองแยกข้อความมีค่าสูงที่สุด ซึ่งสาเหตุได้อภิปรายไว้แล้วในข้อพิจารณาประการที่สอง

นอกจากนี้ เมื่อพิจารณาค่าความสนใจจากการใช้กลไกจุดสนใจตามแบบจำลองที่นำเสนอ สามารถแสดงตัวอย่างได้ตามรูปที่ 5.3

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

การที่	จำเลยร่วม	กับ	พวก	รุม	เตะ	ผู้ตาย
0.0002	0	0.0183	0.0141	0.0052	0.0009	0.0659

ผู้ใด	ฆ่า	ผู้อื่น	ต้อง	ระวางโทษ	ประหารชีวิต
9E-04	0.005	5E-04	5E-04	0.0003	0.0002

จำคุก	ตลอดชีวิต	หรือ	จำคุก	ตั้งแต่	สิบห้า	ปี	ถึง	ยี่สิบ	ปี
2E-04	0.0498	0.065	0.008	9E-04	0.375	0	0.028	7E-04	0.36

รูปที่ 5.3 *ภาพแสดงค่าความสนใจของคำต่าง ๆ จากการใช้กลไกจุดสนใจในประเด็นปัญหาทดสอบหนึ่ง ค่าความสนใจของคำใดมากจะมีพื้นหลังเป็นสีเขียว และค่าที่น้อยลงจะแสดงเป็นสีเหลือง สีส้ม และสีแดงตามลำดับ อนึ่ง ตารางบนสุด คือ คำทั้งหมดในข้อความข้อเท็จจริง ส่วนตารางที่เหลือ คือ คำทั้งหมดในข้อความข้อกฎหมาย*

จากตัวอย่างในรูปที่ 5.3 พบว่า ค่าความสนใจที่มากที่สุดอยู่ที่คำว่า “สิบห้า” มีค่า 0.375 รองลงมา คือ คำว่า “ปี” มีค่า 0.36 ส่วนคำที่ให้ค่าน้ำหนักน้อยที่สุดคือคำว่า “จำเลยร่วม” มีค่าเป็น 0 หรือ ไม่ให้ความสำคัญเลย อย่างไรก็ตาม ค่าน้ำหนักจากกลไกจุดสนใจสามารถสื่อถึงเหตุการณ์ที่เป็นข้อเท็จจริง ตลอดจนฐานความผิดและอัตราโทษได้ เพราะ

1) หากพิจารณาค่าความสนใจที่มีพื้นหลังเป็นสีเขียวและสีเหลืองแล้ว ในข้อความข้อเท็จจริง กลไกจุดสนใจให้น้ำหนักความสนใจถึงลักษณะการกระทำความผิดของจำเลยเอาไว้ คือ การร่วมกับพวก รุมเตะผู้ตาย ส่วนในข้อความข้อกฎหมาย กลไกจุดสนใจให้ความสำคัญกับองค์ประกอบภายนอกของความผิดฐานฆ่าผู้อื่น นั่นคือ การฆ่า ตลอดจนให้ความสำคัญกับอัตราโทษที่จะได้รับในส่วนที่เป็นโทษจำคุกตลอดชีวิต และจำคุกตั้งแต่สิบห้าปีถึงยี่สิบปี
2) ประเด็นปัญหาทดสอบนี้เป็นข้อเท็จจริงในความผิดฐานฆ่าผู้อื่น ซึ่งศาลตัดสินว่าจำเลยมีความผิดฐานฆ่าผู้อื่นโดยเจตนาตามประมวลกฎหมายอาญา มาตรา 288 และแบบจำลองสามารถทำนายผลคำตัดสินได้อย่างถูกต้อง

ข้อพิจารณาประการที่สี่ คือ การใช้กลไกจุดสนใจแบบที่นำเสนอมีประสิทธิภาพดีกว่าการใช้กลไกจุดสนใจในตัวเองแยกข้อความ เพราะค่า $F1$ ที่ได้จากแบบจำลองที่นำเสนอในภาพรวมมีค่าสูงกว่า และผลต่างของ Acc ที่น้อยกว่า แสดงให้เห็นว่าประสิทธิภาพในการทำนายของแบบจำลองที่นำเสนอทั้งที่ทำนายว่า “ผิด” และ “ไม่ผิด” มีคุณภาพใกล้เคียงกัน ไม่โน้มเอียงไปด้านใดด้านหนึ่ง ตลอดจน ไม่ เกิดปัญหาตัวแทนข้อความอ่อนไหวต่อแบบจำลอง

ข้อพิจารณาประการสุดท้าย คือ การใช้โครงข่ายประสาทเทียมแบบโมดูลเพื่อทำนายผลคำตัดสินโดยใช้ข้อมูลขาเข้าเป็นประเด็นในคดีอาญามีประสิทธิภาพดีกว่าการใช้กฎตรรกะประมวลผล เพราะการใช้กฎตรรกะทำให้ค่า $F1$ ของทุกแบบจำลองเทคนิคการเรียนรู้เชิงลึกมีค่าลดลง ดังจะเห็นได้ว่า

- แบบจำลองไม่มีกลไกจุดสนใจพ่วงกฎตรรกะ ทำให้ค่า $F1$ ต่ำลง 7.31% และค่า Acc ต่ำลง 8.13%
- แบบจำลองจุดสนใจในตัวเองแยกข้อความพ่วงกฎตรรกะ ทำให้ค่า $F1$ ต่ำลงถึง 10.38% และค่า Acc ต่ำลง 8.13% เท่ากับแบบจำลองไม่มีกลไกจุดสนใจพ่วงกฎตรรกะ
- แบบจำลองที่นำเสนอพ่วงกฎตรรกะ ทำให้ค่า $F1$ ต่ำลง 8.58% และค่า Acc ต่ำลง 9.35%

ทั้งนี้ สาเหตุที่การใช้กฎตรรกะทำให้ประสิทธิภาพของแบบจำลองลดลงนั้น เนื่องจากกฎตรรกะ ไม่ยืดหยุ่นต่อข้อมูลขาเข้า กล่าวคือ เมื่อผลลัพธ์การทำนายประเด็นในคดีอาญามีค่าอย่างใดอย่างหนึ่งระหว่าง 0 กับ 1 แล้ว หากการทำนายประเด็นมีข้อผิดพลาดแม้เพียงประเด็นเดียว ก็อาจทำให้การทำนายผลคำตัดสินผิดไปเลยได้ เช่น ถ้าประเด็นปัญหาทดสอบหนึ่งมีผลคำตัดสินเฉลยว่า ผิด แล้ว การทำนายประเด็นครบองค์ประกอบภายนอก (isexternalelements) จะต้องให้ผลลัพธ์เป็น 1 ด้วย แต่ถ้าทำนายผิดพลาดเป็น 0 แล้ว การทำนายผลคำตัดสินก็จะผิดตามไปด้วยนั่นเอง ในทางกลับกัน หากเราใช้โครงข่ายประสาทเทียมแบบโมดูลในการทำนายผลคำตัดสินแล้ว โครงข่ายดังกล่าวจะให้ความยืดหยุ่นต่อข้อมูลขาเข้ามากกว่า เพราะในระหว่างการสอนแบบจำลอง หากพบว่าทำนายผิดพลาด ยังสามารถปรับค่าน้ำหนักของโครงข่ายเพื่อเพิ่มโอกาสในการทำนายผลคำตัดสินให้ถูกต้องได้นั่นเอง

ด้วยเหตุนี้ โครงข่ายประสาทเทียมแบบโมดูลจึงสามารถลอกเลียนโครงสร้างความรับผิดทางอาญาได้ อนึ่ง มีข้อน่าสังเกตว่า แบบจำลองที่ใช้เทคนิคการเรียนรู้เชิงลึกร่วมกับกฎตรรกะ จะให้ค่า Acc เท่ากันทั้งหมด คือ 55.28% ซึ่งจากการตรวจสอบพบว่า ไม่ได้ เกิดจากการทำนายคำตอบฝั่งเดียวแต่อย่างใด เพราะแบบจำลองสามารถทำนายคำตอบถูกต้องได้ทั้งที่ตอบว่า “ผิด” และ “ไม่ผิด” ดังปรากฏในตารางที่ 5.5

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.3.2 การทำนายประเด็นในคดีอาญา

สำหรับการทำนายประเด็นในคดีอาญาจำนวน 8 ประเด็นตามที่ระบุไว้ในหัวข้อที่ 4.3.4 นั้น ก่อนอื่นขอแสดงสถิติข้อมูลของแต่ละประเด็นซึ่งปรากฏในชุดข้อมูลสอนและทดสอบ โดยแสดงอยู่ในตารางที่ 5.6 และตารางที่ 5.7 ตามลำดับ

ตารางที่ 5.6 *ตารางแสดงสถิติข้อมูลส่วนประเด็นในคดีอาญาในชุดข้อมูลสอน*

ประเด็น	จำนวนประเด็นปัญหาที่มีคำตอบ "ใช่"	จำนวนประเด็นปัญหาที่มีคำตอบ "ไม่ใช่"	จำนวนประเด็นปัญหาที่ไม่สนใจ	อัตราส่วนใช่ กับ ไม่ใช่
การกระทำ	686	8	0	99 : 1
ครบองค์ประกอบภายนอก	496	190	8	72 : 28
เจตนา	447	27	220	94 : 6
ประมาท	17	5	672	77 : 23
ยกเว้นความผิด	129	333	232	28 : 72
ยกเว้นโทษ	9	324	361	3 : 97
พยายาม	47	260	387	15 : 85
พยายามเป็นไปไม่ได้	7	40	647	15 : 85

ตารางที่ 5.7 *ตารางแสดงสถิติข้อมูลส่วนประเด็นในคดีอาญาในชุดข้อมูลทดสอบ*

ประเด็น	จำนวนประเด็นปัญหาที่มีคำตอบ "ใช่"	จำนวนประเด็นปัญหาที่มีคำตอบ "ไม่ใช่"	จำนวนประเด็นปัญหาที่ไม่สนใจ	อัตราส่วนใช่ กับ ไม่ใช่
การกระทำ	245	1	0	100 : 0
ครบองค์ประกอบภายนอก	170	75	1	69 : 31
เจตนา	162	3	81	98 : 2
ประมาท	4	1	241	80 : 20
ยกเว้นความผิด	16	150	80	10 : 90
ยกเว้นโทษ	1	149	96	1 : 99
พยายาม	27	117	102	19 : 81
พยายามเป็นไปไม่ได้	3	24	219	11 : 89

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ทั้งนี้ การทดสอบประสิทธิภาพการทำนายประเด็นในคดีอาญา มีข้อแตกต่างจากการทดสอบการทำนายผลคำตัดสินอยู่บ้าง ดังต่อไปนี้

1) แบบจำลองจากเทคนิคการเรียนรู้เชิงลึกที่ใช้ในการทำนายประเด็น เป็นแบบจำลองเดียวกันกับที่ใช้ทำนายผลคำตัดสินนั่นเอง เพราะผลลัพธ์การทำนายประเด็นมาจากโครงข่ายประสาทเทียมแบบโมดูล ซึ่งนำไปใช้ทำนายผลคำตัดสินในคดีอาญาด้วยตามโครงสร้างความรับผิดทางอาญา
2) เนื่องด้วยโครงข่ายประสาทเทียมแบบโมดูลมีลักษณะเป็นแบบจำลองร่วม ตามที่กล่าวมาแล้วในหัวข้อที่ 4.3.4 ดังนั้น แบบจำลองที่ใช้เทคนิคการเรียนรู้ของเครื่องเดิมจึง ไม่สามารถ ใช้วิธีการเรียนรู้ประเด็นในคดีอาญาด้วยแบบจำลองร่วมได้ ดังนั้น การทดลองส่วนนี้จึงต้องสร้างแบบจำลองแยกกันเท่าจำนวนประเด็นที่สนใจ และสอนข้อมูลแยกต่างหากจากกัน
3) แบบจำลองที่ใช้เทคนิคการเรียนรู้เชิงลึกร่วมกับกฎตรรกะ ไม่ใช้ ในการทดลองนี้ เนื่องจากแบบจำลองดังกล่าวใช้ทดสอบสมมติฐานข้อ 5 ซึ่งเกี่ยวข้องกับการทำนายผลคำตัดสินเพียงอย่างเดียว
4) เนื่องด้วยบางประเด็นในชุดข้อมูลทีเอสซีซีอาจระบุคำตอบของการทำนายประเด็นนั้นไว้ว่า -1 แต่เพื่อไม่ให้ กระทบกระเทือนต่อการจำแนกประเภทแบบสองคำตอบ ประกอบกับแบบจำลองนี้มีส่วนสำคัญอยู่ที่การทำนายผลคำตัดสิน ซึ่งกฎตรรกะตามโครงสร้างความรับผิดทางอาญาโดยส่วนใหญ่จะประกอบด้วยตัวดำเนินการ “และ” ดังนั้น จึงใช้วิธีการเปลี่ยนคำตอบของประเด็นที่ระบุไว้ว่า -1 เดิมให้กลายเป็น 0

สำหรับผลการทดลองในหัวข้อนี้เป็นไปตามตารางที่ 5.8 และสามารถพิจารณาผลการทดลองที่ตัดคำตอบ -1 ออกไปได้ในตารางที่ 5.9 โดยพบว่า

ประเด็นมีการกระทำหรือไม่ แบบจำลองทุกแบบ ยกเว้น แบบจำลองไม่มีกลไกจุดสนใจ ให้ประสิทธิภาพสูงสุดเท่ากัน โดยมีค่า Acc อยู่ที่ 99.59% และค่า $F1$ อยู่ที่ 49.90% ตามลำดับ และเนื่องจากชุดข้อมูลสอนและทดสอบ ไม่มี คำตอบที่ตอบว่า -1 เลย ผลการทดลองทั้งสองตารางในส่วนนี้จึงเหมือนกัน สำหรับสาเหตุที่แบบจำลองแทบทุกแบบมีประสิทธิภาพสูงเช่นนี้เกิดจากชุดข้อมูลที่ใช้สอนและทดสอบ ไม่ต่ำกว่าร้อยละ 99 ล้วนให้คำตอบว่าข้อเท็จจริงนี้ มี การกระทำตามโครงสร้างความรับผิดทางอาญาทั้งสิ้น ซึ่งสอดคล้องกับคดีอาญาส่วนใหญ่ที่จำเลยมักมีการกระทำเสมอ กรณี ไม่มี การกระทำ เช่น การเคลื่อนไหวร่างกายเพราะละเมอหรือเป็นโรคลมชัก ฯลฯ อนึ่ง เหตุที่แบบจำลองไม่มีกลไกจุดสนใจได้ประสิทธิภาพ ไม่เท่ากับ แบบจำลองอื่น เกิดจากแบบจำลองนี้พยายามตอบคำตอบเป็น 0 (ไม่มีการกระทำ) ด้วย แต่ยังไม่แม่นยำพอ

ประเด็นการกระทำครบองค์ประกอบภายนอกของความผิดหรือไม่ แบบจำลองที่นำเสนอให้ประสิทธิภาพสูงสุดด้วยค่า $F1$ อยู่ที่ 60.30% แต่เอสวีเอ็มให้ค่า Acc สูงสุดอยู่ที่ 70.33% นอกจากนี้ หากพิจารณาโดยตัดผลลัพธ์ -1 ออกไป พบว่าแบบจำลองที่นำเสนอให้ประสิทธิภาพสูงสุดด้วยค่า $F1$ อยู่ที่ 59.94% แต่เอสวีเอ็มให้ค่า Acc สูงสุดอยู่ที่ 70.20% แสดงว่าแบบจำลองที่นำเสนอสามารถทำนายผลลัพธ์ได้โดยไม่เอนเอียง ไปข้างใดข้างหนึ่ง แตกต่างจากเอสวีเอ็ม ซึ่งสังเกตได้ว่าค่า $F1$ นั้น ไม่ถึง 50%

ตารางที่ 5.8 *ตารางแสดงผลการทดลองวัดประสิทธิภาพการทำนายประเด็นในคดีอาญาจากแบบจำลองต่าง ๆ*

*โดยถือว่าประเด็นที่มีผลลัพธ์ -1 มีค่าเป็น 0 ผลที่ดีที่สุดจะแสดงเป็น**ตัวหนาและขีดเส้นใต้***

แบบจำลอง	มีการกระทำหรือไม่		ครบองค์ประกอบภายนอก		เจตนาหรือไม่		ประมาทหรือไม่	
	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)
เนอีฟเบยส์	**99.59**	**49.90**	69.11	40.87	65.85	39.71	98.37	49.59
เอสวีเอ็ม	**99.59**	**49.90**	**70.33**	49.93	**70.33**	56.98	98.37	66.26
ไม่มีกลไกจุดสนใจ	97.56	49.38	64.63	60.09	58.94	56.61	98.78	78.26
จุดสนใจในตัวเองแยกข้อความ	**99.59**	**49.90**	65.45	57.01	59.76	55.13	**99.19**	**83.13**
แบบจำลองที่นำเสนอ	**99.59**	**49.90**	65.85	**60.30**	61.38	**59.35**	**99.19**	**83.13**

แบบจำลอง	ยกเว้น ความผิด หรือไม่		ยกเว้น โทษ หรือไม่		พยายามหรือไม่		พยายามเป็นไปไม่ได้หรือไม่	
	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)
เนอีฟเบยส์	93.50	48.32	**99.59**	**49.90**	89.02	47.10	98.78	49.69
เอสวีเอ็ม	95.53	83.93	**99.59**	**49.90**	**94.31**	**85.44**	**99.59**	**89.90**
ไม่มีกลไกจุดสนใจ	**96.34**	**86.16**	98.78	49.69	87.80	70.68	97.56	49.38
จุดสนใจในตัวเองแยกข้อความ	94.72	72.61	99.19	49.80	91.46	69.31	98.78	49.69
แบบจำลองที่นำเสนอ	93.09	75.16	**99.59**	**49.90**	91.46	70.76	98.37	49.59

ตารางที่ 5.9 *ตารางแสดงผลการทดลองวัดประสิทธิภาพการทำนายประเด็นในคดีอาญาจากแบบจำลองต่าง ๆ โดยตัดประเด็นที่มีผลลัพธ์เป็น -1 ออกไป ผลที่ดีที่สุดจะแสดงเป็น* ***ตัวหนาและขีดเส้นใต้***

แบบจำลอง	มีการกระทำหรือไม่		ครบองค์ประกอบภายนอก		เจตนาหรือไม่		ประมาทหรือไม่	
	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)
เนอีฟเบยส์	**99.59**	**49.90**	69.39	40.96	**98.18**	**49.54**	20.00	16.67
เอสวีเอ็ม	**99.59**	**49.90**	**70.20**	48.99	93.94	48.44	20.00	16.67
ไม่มีกลไกจุดสนใจ	97.56	49.38	64.49	59.76	62.42	41.29	**60.00**	**58.33**
จุดสนใจในตัวเองแยกข้อความ	**99.59**	**49.90**	65.71	57.25	69.09	42.68	**60.00**	**58.33**
แบบจำลองที่นำเสนอ	**99.59**	**49.90**	65.71	**59.94**	63.03	40.16	**60.00**	**58.33**

แบบจำลอง	ยกเว้น ความผิด หรือไม่		ยกเว้น โทษ หรือไม่		พยายามหรือไม่		พยายามเป็นไปไม่ได้หรือไม่	
	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)
เนอีฟเบยส์	90.36	47.47	**99.33**	**49.83**	81.25	44.83	88.89	47.06
เอสวีเอ็ม	94.58	85.63	**99.33**	**49.83**	**93.06**	**87.90**	**96.30**	**88.98**
ไม่มีกลไกจุดสนใจ	**95.78**	**88.22**	98.00	49.49	79.17	67.62	77.78	43.75
จุดสนใจในตัวเองแยกข้อความ	92.77	73.05	**99.33**	**49.83**	85.42	67.44	88.89	47.06
แบบจำลองที่นำเสนอ	91.57	77.06	**99.33**	**49.83**	86.11	69.68	85.19	46.00

ประเด็นจำเลยมีเจตนากระทำความผิดหรือไม่ แบบจำลองที่นำเสนอให้ประสิทธิภาพสูงสุดด้วยค่า $F1$ อยู่ที่ 59.35% แต่เอสวีเอ็มให้ค่า Acc สูงสุดอยู่ที่ 70.33% กรณีนี้คล้ายคลึงกับประเด็นการกระทำครบองค์ประกอบภายนอกของความผิดหรือไม่ซึ่งได้เอ่ยถึงไปแล้ว ทว่าหากพิจารณาผลการทดลองโดยตัดผลลัพธ์ -1 ออกไป พบว่าเนอีฟเบยส์ให้ประสิทธิภาพสูงสุดด้วยค่า $F1$ 49.54% แต่ค่า Acc สูงถึง 98.18% เมื่อตรวจสอบในรายละเอียดพบว่าเนอีฟเบยส์ทำนายผลลัพธ์ทุกค่าเป็น 1 ทั้งหมด (เป็นการกระทำโดยเจตนา) นอกจากนี้หากพิจารณาจากสัดส่วนข้อมูลสอนในประเด็นนี้ พบว่ามีจำนวนคำตอบเป็น 1 สูงถึงร้อยละ 94 ในขณะที่สัดส่วนข้อมูลทดสอบ พบว่ามีจำนวนคำตอบว่า 1 สูงถึงร้อยละ 98 ทำให้เนอีฟเบยส์โน้มเอียงในการทำนายนั่นเอง ดังนั้นหากตัดเนอีฟเบยส์ออกไป แบบจำลองที่ให้ประสิทธิภาพสูงสุดจะเป็นเอสวีเอ็ม

ประเด็นจำเลยกระทำความผิดโดยประมาทหรือไม่ แบบจำลองที่นำเสนอและแบบจำลองจุดสนใจในตัวเองแยกข้อความให้ประสิทธิภาพสูงที่สุด ด้วยค่า Acc อยู่ที่ 99.19% และค่า $F1$ อยู่ที่ 83.13% ขณะเดียวกันเมื่อพิจารณาผลลัพธ์หลังตัด -1 ออกไป พบว่าแบบจำลองทั้งสองรวมถึงแบบจำลองไม่มีกลไกจุดสนใจให้ประสิทธิภาพสูงที่สุด ด้วยค่า Acc อยู่ที่ 60.00% และค่า $F1$ อยู่ที่ 58.33% ทั้งนี้ เมื่อตรวจสอบสัดส่วนข้อมูลสอนและทดสอบของประเด็นนี้แล้ว พบว่าชุดข้อมูลสอนในประเด็นประมาทมีข้อมูลที่ให้คำตอบ 1 กับ 0 รวมกันเพียง 22 ประเด็นปัญหาเท่านั้น ที่เหลือล้วนมีค่าเป็น -1 ส่วนชุดข้อมูลทดสอบมีข้อมูลที่ให้คำตอบ 1 กับ 0 รวมกันเพียง 5 ประเด็นปัญหา ดังนั้น แสดงว่าข้อมูลคำพิพากษาศาลฎีกาที่ผู้วิจัยรวบรวมมาในความผิดที่ได้กระทำโดยประมาท และสามารถทำนายประเด็นปัญหานี้ได้ยังมีน้อยมาก หากรวบรวมเพิ่มขึ้นมาอีกได้ในอนาคตอาจช่วยให้การทำนายประเด็นนี้สมบูรณ์มากยิ่งขึ้น

ประเด็นมีเหตุยกเว้น ความผิด หรือไม่ แบบจำลองไม่มีกลไกจุดสนใจให้ประสิทธิภาพสูงที่สุด ด้วยค่า Acc อยู่ที่ 96.34% และค่า $F1$ อยู่ที่ 86.16% ขณะเดียวกันเมื่อพิจารณาผลลัพธ์หลังตัด -1 ออกไป พบว่าแบบจำลองข้างต้นยังคงให้ประสิทธิภาพสูงสุดด้วยค่า Acc อยู่ที่ 95.78% และค่า $F1$ อยู่ที่ 88.22% ทั้งนี้ พบว่าแบบจำลองที่ใส่กลไกจุดสนใจเข้าไปกลับมีประสิทธิภาพต่ำกว่าแบบจำลองไม่มีกลไกจุดสนใจ ซึ่งเมื่อตรวจสอบจำนวนคำตอบที่แบบจำลองดังกล่าวทำนายออกมา พบว่าแบบจำลองที่ใส่กลไกจุดสนใจพยายามทำนายผลลัพธ์ค่า 1 ซึ่งมีสัดส่วนข้างน้อยทั้งชุดข้อมูลสอนและทดสอบให้มากขึ้นกว่าแบบจำลองไม่มีกลไกจุดสนใจ ทว่าความแม่นยำยังไม่สูงพอ ทำให้ประสิทธิภาพตกลงไปนั่นเอง

ประเด็นมีเหตุยกเว้น โทษ หรือไม่ แบบจำลองที่นำเสนอ ร่วมด้วยเนอีฟเบยส์และเอสวีเอ็มให้ประสิทธิภาพสูงที่สุด ด้วยค่า Acc อยู่ที่ 99.59% และค่า $F1$ อยู่ที่ 49.90% ขณะเดียวกันเมื่อพิจารณาผลลัพธ์หลังตัด -1 ออกไป พบว่าแบบจำลองทุกแบบ ยกเว้นแบบจำลองไม่มีกลไกจุดสนใจ ให้ประสิทธิภาพสูงสุดด้วยค่า Acc อยู่ที่ 99.33% และค่า $F1$ อยู่ที่ 49.83% ตามลำดับ สาเหตุของปรากฏการณ์นี้เป็นไปทำนองเดียวกันกับประเด็นมีการกระทำหรือไม่ เนื่องจากชุดข้อมูลที่ใช้สอนและทดสอบ ไม่ต่ำกว่าร้อยละ 97 ล้วนให้คำตอบว่า

ข้อเท็จจริงนี้ ไม่มี เหตุยกเว้นโทษ ทั้งสิ้น นอกจากนี้ เหตุยกเว้นโทษที่ปรากฏในชุดข้อมูลทีเอสซีซี โดยหลักแล้วจะมาจากการพิสูจน์ความจริงในความผิดฐานหมิ่นประมาทตามประมวลกฎหมายอาญา มาตรา 330

ประเด็นพยายามกระทำความผิดหรือไม่ พบว่า เอสวีเอ็มให้ประสิทธิภาพสูงที่สุด ด้วยค่า Acc อยู่ที่ 94.31% และค่า $F1$ อยู่ที่ 85.44% นอกจากนี้ หากพิจารณาโดยตัดผลลัพธ์ที่มีค่า -1 ออกไป จะพบว่าเอสวีเอ็มยังคงให้ประสิทธิภาพสูงที่สุดเช่นเดิม ด้วยค่า Acc อยู่ที่ 93.06% และค่า $F1$ อยู่ที่ 87.90% อย่างไรก็ตาม หากไม่นับเอสวีเอ็มแล้ว จะพบว่าแบบจำลองที่นำเสนอมีประสิทธิภาพสูงที่สุดทั้งกรณีที่นับรวมและไม่นับรวม -1

ประเด็นพยายามกระทำความผิดซึ่งเป็นไปไม่ได้โดยแน่แท้หรือไม่ พบว่า เอสวีเอ็มให้ประสิทธิภาพสูงที่สุด ด้วยค่า Acc อยู่ที่ 99.59% และค่า $F1$ อยู่ที่ 89.90% นอกจากนี้ หากพิจารณาโดยตัดผลลัพธ์ที่มีค่า -1 ออกไปแล้ว จะพบว่าเอสวีเอ็มยังคงให้ประสิทธิภาพสูงที่สุดเช่นเดิม ด้วยค่า Acc อยู่ที่ 96.30% และค่า $F1$ อยู่ที่ 88.98% แต่ถ้าไม่นับเอสวีเอ็มแล้ว จะพบว่าแบบจำลองที่นำเสนอมีประสิทธิภาพสูงที่สุดทั้งกรณีที่นับรวมและไม่นับรวม -1 เหมือนอย่างกรณีพยายามกระทำความผิดด้วย ณ จุดนี้ สอดคล้องกับสัดส่วนชุดข้อมูลสอน และข้อมูลทดสอบที่มีปริมาณใกล้เคียงกัน ตลอดจนการพยายามกระทำความผิดทั้งสองรูปแบบตามประมวลกฎหมายอาญาใช้หลักเกณฑ์พิจารณาอย่างเดียวกัน มีเพียงกรณีเป็นไปไม่ได้อย่างแน่แท้ซึ่งมีเกณฑ์เพิ่มขึ้นมาอีก 1 ข้อเท่านั้น ด้วยเหตุนี้ ผลลัพธ์การทดลองของประเด็นพยายามกระทำความผิดจึงให้ประสิทธิภาพการทำนายที่ใกล้เคียงกัน

โดยสรุปแล้ว จากการทดลองประสิทธิภาพการทำนายประเด็นในคดีอาญา แบบจำลองที่นำเสนอสามารถให้ประสิทธิภาพสูงสุดในการทำนายประเด็นในคดีอาญา 5 ประเด็นจากทั้งหมด 8 ประเด็น ซึ่งล้วนอยู่ในประเด็นที่นำไปสู่ผลคำตัดสินในคดีอาญาทั้งสิ้น และหากพิจารณากรณีตัดผลลัพธ์ -1 ออกไป แบบจำลองที่นำเสนอให้ประสิทธิภาพสูงสุดในการทำนาย 4 ประเด็น ซึ่งยังคงเป็นประเด็นที่นำไปสู่ผลคำตัดสินเช่นกัน ถึงกระนั้นก็ตาม แบบจำลองที่นำเสนอยังคงต้องผ่านการปรับปรุงเพิ่มเติมต่อไปเพื่อให้ได้ประสิทธิภาพในการทำนายสูงขึ้น ซึ่งหากประสิทธิภาพการทำนายประเด็นในคดีอาญาสูงแล้ว โดยเฉพาะใน 6 ประเด็นแรก เป็นที่คาดหมายได้ว่าประสิทธิภาพในการทำนายผลคำตัดสินในคดีอาญาจะสูงขึ้นตามไปด้วยในที่สุด

อนึ่ง มีข้อสังเกตถึงกรณีที่เอสวีเอ็มให้ค่า Acc สูงสุดในการทำนายประเด็น องค์ประกอบภายนอกของความผิด และประเด็น เจตนา ตลอดจนให้ค่า Acc และ $F1$ สูงสุดในประเด็น พยายามกระทำความผิด นั้นสาเหตุเกิดจากแบบจำลองเทคนิคการเรียนรู้เชิงลึกเป็นแบบจำลองร่วมซึ่งรวมค่าความผิดพลาดทั้งหมดจากการทำนายผลคำตัดสินและประเด็นในคดีอาญาแล้วนำมาใช้เป็นฟังก์ชันการสูญเสียสำหรับสอนแบบจำลอง ซึ่งค่าความผิดพลาดจากแต่ละส่วน ไม่ได้ ถ่วงน้ำหนักเอาไว้ ในขณะที่เอสวีเอ็มจะสอนแบบจำลองโดย ไม่ได้ อ้างอิงค่าความผิดพลาดของการทำนายผลคำตัดสินและประเด็นอื่น ๆ ร่วมด้วยเลย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 6
สรุปผลการทดลองและแนวทางการวิจัยในขั้นถัดไป

6.1 สรุปผลการทดลอง

วิทยานิพนธ์ฉบับนี้ นำเสนอแบบจำลองทำนายผลคำตัดสินและประเด็นในคดีอาญาที่พัฒนาขึ้นด้วยเทคนิคการเรียนรู้เชิงลึก โดยเรียนรู้ข้อมูลจากคำพิพากษาศาลฎีกาในคดีอาญา ทั้งนี้ โครงสร้างภายในแบบจำลองเริ่มต้นจากการรับข้อความข้อเท็จจริงและข้อกฎหมายที่ตัดคำแล้วไปผ่านตารางค้นหาเพื่อฝังคำ จากนั้นนำไปผ่านโครงข่ายประตูวกกลับสองทิศทางให้ได้เวกเตอร์ข้อความของข้อเท็จจริงและข้อกฎหมาย เวกเตอร์ทั้งสองที่ได้จะผ่านการพิจารณากลไกจุดสนใจในตัวเองร่วมกันเพื่อค้นหาคำที่สำคัญจากทั้งข้อเท็จจริงและข้อกฎหมาย ก่อนนำไปผ่านโครงข่ายประสาทเทียมแบบโมดูลเพื่อทำนายผลคำตัดสินและประเด็นซึ่งออกแบบตามโครงสร้างความรับผิดทางอาญา

ผลการทดลองพบว่า แบบจำลองนี้ให้ประสิทธิภาพดีกว่าแบบจำลองที่ใช้เทคนิคการเรียนรู้ของเครื่องเดิม โดยในการทำนายผลคำตัดสินจากชุดข้อมูลทดสอบนั้น แบบจำลองให้ค่าความแม่นยำอยู่ที่ 64.63% และค่าเฉลี่ยมหภาคของเอฟวันอยู่ที่ 63.82% ส่วนการทำนายประเด็นในคดีอาญา พบว่าแบบจำลองที่นำเสนอให้ประสิทธิภาพการทำนายสูงกว่าแบบจำลองที่ใช้เทคนิคการเรียนรู้ของเครื่องเดิมใน 5 ประเด็นสำคัญที่มีผลต่อคำตัดสินในคดีอาญา นอกจากนี้ มีข้อสังเกตเพิ่มเติมว่า

1) การใช้เวกเตอร์จากโครงข่ายประตูวกกลับสองทิศทางเป็นตัวแทนข้อความช่วยให้ประสิทธิภาพในการทำนายผลคำตัดสินสูงขึ้นเมื่อเทียบกับการใช้ตัวแทนข้อความรูปแบบดั้งเดิม นอกจากนี้ ยังช่วยป้องกันปัญหาตัวแทนข้อความอ่อนไหวต่อแบบจำลองด้วย แต่ต้อง ไม่ใส่ กลไกจุดสนใจในตัวเองแยกข้อความ
2) การใช้กลไกจุดสนใจช่วยให้ประสิทธิภาพการทำนายของแบบจำลองดีขึ้น
3) การใช้กลไกจุดสนใจแบบที่นำเสนอมีประสิทธิภาพดีกว่าการใช้กลไกจุดสนใจในตัวเองแยกข้อความ รวมถึงทนทานปัญหาตัวแทนข้อความอ่อนไหวต่อแบบจำลองด้วย
4) การใช้โครงข่ายประสาทเทียมแบบโมดูลเพื่อทำนายผลคำตัดสินโดยใช้ข้อมูลขาเข้าเป็นประเด็นในคดีอาญาสามารถลอกเลียนโครงสร้างความรับผิดทางอาญาได้ สะท้อนจากประสิทธิภาพการทำนายที่ดีกว่าการใช้กฎตรรกะประมวลผล

อย่างไรก็ดี เนื่องจากประสิทธิภาพการทำนายผลคำตัดสินของแบบจำลองที่นำเสนอยังมีค่าสูงกว่าแบบจำลองตั้งต้นไม่มากนัก ประกอบกับประสิทธิภาพการทำนายประเด็นในคดีอาญายัง ไม่สามารถ มีประสิทธิภาพสูงสุดได้ในทุกประเด็นที่สนใจ ดังนั้น แบบจำลองที่นำเสนอจึงยังต้องปรับปรุงและพัฒนาต่อไปเพื่อให้มีประสิทธิภาพสูงขึ้น ผ่านการวิจัยในอนาคต

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

6.2 แนวทางการวิจัยในขั้นถัดไป

1) การเพิ่มปริมาณข้อมูลในชุดข้อมูลทีเอสซีซีเพื่อช่วยยืนยันประสิทธิภาพของแบบจำลอง เนื่องจากแบบจำลองที่สร้างขึ้นด้วยเทคนิคการเรียนรู้เชิงลึกจะมีประสิทธิภาพมากเมื่อสอนแบบจำลองกับข้อมูลปริมาณมหาศาล แต่ข้อมูลในชุดข้อมูลนี้ยังมีปริมาณไม่มากนัก โดยจากการทดลอง มีจำนวนคำในชุดข้อมูลทั้งหมด 88,101 คำ
2) การปรับปรุงการทดลองเพื่อแก้ไขปัญหาความ ไม่สมดุล ระหว่างคำตอบของประเด็นในคดีอาญาซึ่งมีที่มาจากชุดข้อมูลทีเอสซีซี
3) การทดลองเปลี่ยนข้อมูลขาเข้าในส่วนข้อกฎหมาย โดยตัดอัตราโทษออกไปเพื่อพิจารณาว่า ข้อความระบุอัตราโทษมีผลต่อประสิทธิภาพของแบบจำลองหรือไม่
4) การทดลองสร้างเวกเตอร์คำจากเวกเตอร์อักขระ (Character Embedding) ซึ่งเป็นวิธีการหนึ่งที่ได้รับความสนใจมากขึ้นในงานประมวลผลภาษาธรรมชาติที่ใช้เทคนิคการเรียนรู้เชิงลึก เนื่องจากแบบจำลองในขณะนี้ยังใช้วิธีการตัดคำในรูปแบบดั้งเดิม แม้จะเพิ่มเติมคำศัพท์ลงไปจนมีถึง 88,130 คำ ก็ยังมีโอกาสพบปัญหาคำไม่รู้จักในคลังศัพท์ (Out-of-vocabulary : OOV) ซึ่งส่งผลต่อความแม่นยำของแบบจำลองได้ โดยในการทดลอง พบว่าจากจำนวนคำในคลังศัพท์ของข้อมูลสอนทั้งหมด 4,317 คำ มีจำนวนคำไม่รู้จักในคลังศัพท์ทั้งหมด 1,714 คำ คิดเป็น 39.70%
5) การทดลองใช้กลไกจุดสนใจในรูปแบบอื่น ๆ เพื่อโอกาสในการเพิ่มประสิทธิภาพของแบบจำลอง เช่น กลไกจุดสนใจร่วม (Co-attention) ตามผลงานวิจัย [25, 26]
6) การปรับปรุงฟังก์ชันการสูญเสียที่ปัจจุบันใช้วิธีรวมค่าความผิดพลาดของการทำนายผลคำตัดสินและการทำนายทุกประเด็นรวมกัน ซึ่งส่งผลต่อประสิทธิภาพในการทำนายผลคำตัดสินและประเด็นต่าง ๆ ทั้งหมด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รายการอ้างอิง

[1] สำนักงานคณะกรรมการกฤษฎีกา, ประมวลกฎหมายอาญา, 2017. [Online]. Available: http://web.krisdika.go.th/data/law/law4/%bb06/%bb06-20-9999-update.pdf.

[2] เกียรติขจร วัจนะสวัสดิ์, คำอธิบายกฎหมายอาญา ภาค 1 บทบัญญัติทั่วไป, 10th (แก้ไขเพิ่มเติม) ed. สำนักพิมพ์ พลสยาม พริ้นติ้ง (ประเทศไทย), 2008, p. 1000.

[3] กิตติศักดิ์ ปรกติ, ความเป็นมาและหลักการใช้นิติวิธีในระบบซิวิลลอว์และคอมมอนลอว์, 5 ed. กรุงเทพมหานคร: วิญญูชน, 2017, p. 100.

[4] สมยศ เชื้อไทย, คำอธิบายวิชากฎหมายแพ่ง : หลักทั่วไป, 23 ed. กรุงเทพมหานคร: วิญญูชน, 2017, p. 220.

[5] ศูนย์เทคโนโลยีสารสนเทศและการสื่อสารในศาลฎีกา. ระบบสืบค้นคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา [Online]. Available: http://deka.supremecourt.or.th/

[6] สำนักงานคณะกรรมการกฤษฎีกา, พระธรรมนูญศาลยุติธรรม, 2015. [Online]. Available: http://web.krisdika.go.th/data/law/law2/%be16/%be16-20-9999-update.pdf.

[7] สำนักงานคณะกรรมการกฤษฎีกา, ประมวลกฎหมายวิธีพิจารณาความอาญา, 2016. [Online]. Available: http://web.krisdika.go.th/data/law/law4/%bb05/%bb05-20-9999-update.pdf.

[8] K. Kowsrihawat and P. Vateekul, "An information extraction framework for legal documents: A case study of Thai Supreme Court verdicts," in Proceedings of the 2015 12th International Joint Conference on Computer Science and Software Engineering, JCSSE 2015, 2015, pp. 275-280.

[9] คณพล จันทน์หอม, รากฐานกฎหมายอาญา. กรุงเทพมหานคร: วิญญูชน, 2015, p. 126.

[10] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A Neural Probabilistic Language Model," Journal of Machine Learning Research, Conference Paper vol. 3, no. 6, pp. 1137-1155, 2003.

[11] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations ofwords and phrases and their compositionality," in Advances in Neural Information Processing Systems, 2013.

[12] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2014, pp. 1532-1543.

[13] T. Mikolov, W. T. Yih, and G. Zweig, "Linguistic regularities in continuous spaceword representations," in NAACL HLT 2013 - 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Main Conference, 2013, pp. 746-751.

[14] T. Mitchell, Machine Learning. McGraw-Hill, 1997, p. 432.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

[15] D. Jurafsky and J. H. Martin, "Neural Nets and Neural Language Models," Speech and Language Processing, August 28, 2017. Accessed on: April 21, 2018Available: https://web.stanford.edu/~jurafsky/slp3/8.pdf

[16] "Multi-Layer Feedforward Neural Network," in https://commons.wikimedia.org/wiki/File:MultiLayerNeuralNetworkBigger_english.png, MultiLayerNeuralNetworkBigger_english.png, Ed., 1 ed: Wikimedia, 2010.

[17] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," Journal of Machine Learning Research, Article vol. 15, pp. 1929-1958, 2014.

[18] Y. Gal and Z. Ghahramani, "A theoretically grounded application of dropout in recurrent neural networks," in Advances in Neural Information Processing Systems, 2016, pp. 1027-1035.

[19] Y. Lecun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436-444, 27 May, 2015 2015.

[20] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[21] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to Forget: Continual Prediction with LSTM," Neural Computation, vol. 12, no. 10, pp. 2451-2471, 2000/10/01 2000.

[22] K. Cho et al., "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1724-1734: Association for Computational Linguistics.

[23] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," ArXiv e-prints, Accessed on: December 01, 2014Available: https://ui.adsabs.harvard.edu/#abs/2014arXiv1412.3555C

[24] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," ArXiv e-prints, Accessed on: September 01, 2014Available: https://ui.adsabs.harvard.edu/#abs/2014arXiv1409.0473B

[25] J. Lu, J. Yang, D. Batra, and D. Parikh, "Hierarchical question-image co-attention for visual question answering," in Advances in Neural Information Processing Systems, 2016, pp. 289-297.

[26] C. Xiong, V. Zhong, and R. Socher, "Dynamic Coattention Networks For Question Answering," ArXiv e-prints, Accessed on: November 01, 2016Available: https://ui.adsabs.harvard.edu/#abs/2016arXiv161101604X

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

[27] N. Aletras, D. Tsarapatsanis, D. Preotiuc-Pietro, and V. Lampos, "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective," PeerJ Computer Science, 2016.

[28] D. M. Katz, M. J. Bommarito, and J. Blackman, "A general approach for predicting the behavior of the Supreme Court of the United States," PLoS ONE, 2017.

[29] H. J. Spaeth, L. Epstein, A. D. Martin, J. A. Segal, T. J. Ruger, and S. C. Bennesh. Supreme Court Database [Online]. Available: http://scdb.wustl.edu

[30] S. Thammaboosadee and U. Silparcha, "A framework for criminal judicial reasoning system using data mining techniques," in 2008 2nd IEEE International Conference on Digital Ecosystems and Technologies, IEEE-DEST 2008, 2008, pp. 518-523.

[31] S. Thammaboosadee, B. Watanapa, J. H. Chan, and U. Silparcha, "A Two-Stage Classifier That Identifies Charge and Punishment under Criminal Law of Civil Law System," IEICE TRANSACTIONS on Information and Systems, vol. E97-D, no. 4, pp. 864-875, 2014.

[32] P. Osathitporn, N. Soonthornphisaj, and W. Vatanawood, "A scheme of criminal law knowledge acquisition using ontology," in Proceedings - 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2017, 2017, pp. 29-34.

[33] T. Tantisripreecha and N. Soonthornphisaj, "LegalEX: An expert system for law firm," Intelligent Decision Technologies, vol. 10, pp. 315-328, 2016.

[34] ศูนย์วิจัยเทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. โปรแกรมตัดคำสำหรับข้อความภาษาไทย เล็กซ์โต (Thai Lexeme Tokenizer : LexTo) [Online]. Available: http://www.sansarn.com/lexto/

[35] Catmium. PyLexTo [Online]. Available: https://github.com/catmium/PyLexTo

[36] ราชบัณฑิตยสถาน, "พจนานุกรมศัพท์กฎหมายไทย ฉบับราชบัณฑิตยสถาน," 4 ed. กรุงเทพมหานคร, 2013.

[37] Y. Shen, P.-S. Huang, J. Gao, and W. Chen, "ReasoNet: Learning to Stop Reading in Machine Comprehension," eprint arXiv:1609.05284, p. arXiv:1609.05284, 2016.

[38] M. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi, "Bidirectional Attention Flow for Machine Comprehension," eprint arXiv:1611.01603, p. arXiv:1611.01603, 2016.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ก.

ตัวอย่างแถวข้อมูลในตารางต่าง ๆ ของชุดข้อมูลทีเอสซีซี

ตาราง คำพิพากษา

ลำดับที่	ชื่อคุณลักษณะ	ตัวอย่างข้อมูล	คำอธิบาย
1	issueid	1	หมายเลขไอดีของประเด็นปัญหา
2	dekaid	1478/2528	หมายเลขคำพิพากษาศาลฎีกา
3	year	2528	ปีที่พิพากษาคดี
4	category	LB	หมวดความผิด ในที่นี้ คือ หมวดชีวิตและร่างกาย
5	issueno	1	ประเด็นปัญหาที่สกัดได้จากคำพิพากษาศาลฎีกา ลำดับที่ 1
6	lawids	CC-288-00,CC-083-00, CC-063-00	รหัสข้อกฎหมายซึ่งเป็นคีย์หลักของแถวในตาราง ข้อกฎหมาย ในที่นี้ คือ ประมวลกฎหมายอาญา มาตรา 288 ประกอบมาตรา 83 และ 63
7	fact	จำเลยกับพวกร่วมกันใช้อาวุธปืนยิงผู้ตายถูกที่ด้านหลัง กระสุนปืนตัดบริเวณไขสันหลังขาด ผู้ตายเป็นอัมพาตตั้งแต่เอวจนจดเท้าและถึงแก่ความตายสืบเนื่องมาจากบาดแผลที่ถูกยิงและภาวะติดเชื้ออย่างรุนแรงหลังจากเกิดเหตุ 9 เดือนเศษ <discr>ดังนี้ ผู้ตายถึงแก่ความตายสืบเนื่องมาจากบาดแผลที่ถูกยิง </discr> แม้จะเนื่องจากการรักษาไม่ดีจนบาดแผลติดเชื้อ <discr>ก็เป็นผลธรรมดาอันสืบเนื่องมาจากการกระทำของจำเลย	ข้อความ ข้อเท็จจริง ที่สกัดได้จากย่อสั้นของคำพิพากษา ส่วนที่อยู่ภายใต้แท็ก <discr> คือ ดุลพินิจของศาลในคดีนั้น ซึ่งจะนำไปสู่คำตัดสินต่อไป

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ลำดับที่	ชื่อคุณลักษณะ	ตัวอย่างข้อมูล	คำอธิบาย
		</discr>	
8	decision	จำเลยจึงมีความผิดฐานฆ่าผู้ตายโดยเจตนา	ข้อความ คำตัดสิน ที่สกัดได้จากย่อสั้นของคำพิพากษา
9	isact	1	จำเลยมีการกระทำ
10	isexternalelements	1	การกระทำของจำเลยครบองค์ประกอบ *ภายนอก* ของความผิด
11	isinternalelements	1	การกระทำของจำเลยครบองค์ประกอบ *ภายใน* ของความผิด
12	isintent	1	การกระทำของจำเลยเป็นการกระทำโดย *เจตนา*
13	isneglect	-1	ไม่สนใจว่า การกระทำของจำเลยเป็นการกระทำโดย *ประมาท* หรือไม่ เพราะเจตนาและประมาทไม่เกิดร่วมกัน
14	iscause	1	การกระทำของจำเลยสัมพันธ์กับผลแห่งการกระทำ
15	isjustify	0	การกระทำของจำเลย ไม่มี เหตุยกเว้น*ความผิด*
16	isexcuse	0	การกระทำของจำเลย ไม่มี เหตุยกเว้น*โทษ*
17	isguilty	1	จำเลยมีความผิดและต้องรับโทษ
18	isattempt	0	จำเลย ไม่ได้ พยายามกระทำความผิดเพราะความผิดสำเร็จแล้ว
19	isatttemptimpossible	0	จำเลย ไม่ได้ พยายามกระทำความผิดซึ่งเป็นไปไม่ได้โดยแน่แท้ เพราะความผิดสำเร็จแล้ว

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง ข้อกฎหมาย

กรณีที่ 1: ข้อกฎหมายที่มีมาตั้งแต่เริ่มใช้ประมวลกฎหมายอาญา เมื่อวันที่ 1 มกราคม พ.ศ.2500

ลำดับที่	ชื่อคุณลักษณะ	ตัวอย่าง	คำอธิบาย
1	lawid	CC-288-00	หมายเลขไอดีของข้อกฎหมาย มีที่มาจาก lawcode ต่อด้วย lawsection ต่อด้วยจำนวนครั้งในการแก้ไข ถ้าเป็นข้อกฎหมายที่มีขึ้นเป็นครั้งแรก จะเริ่มด้วย 00
2	lawcode	CC	ประมวลกฎหมายอาญา
3	lawsection	288	มาตรา 288
4	category	2-10-1	ภาค 2 ลักษณะ 10 หมวด 1 แห่งประมวลกฎหมายอาญา
5	effectivedate	2500-01-01	วันที่ข้อกฎหมายนี้มีผลบังคับใช้ นั่นคือ วันที่ 1 มกราคม พ.ศ.2500
6	enddate	0000-00-00	วันที่ข้อกฎหมายนี้สิ้นสุดการบังคับใช้ ในที่นี้ หมายความว่าข้อกฎหมายนี้ยังใช้บังคับอยู่จนถึงปัจจุบัน
7	editreference		กฎหมายที่แก้ไขเพิ่มเติมข้อกฎหมายนี้ ในที่นี้ เว้นว่าง เพราะข้อกฎหมายนี้มีมาตั้งแต่เริ่มใช้ประมวลกฎหมายอาญาแล้ว
8	content	ผู้ใดฆ่าผู้อื่น ต้องระวางโทษประหารชีวิต จำคุกตลอดชีวิต หรือจำคุกตั้งแต่สิบห้าปีถึงยี่สิบปี	เนื้อหาข้อกฎหมาย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

กรณีที่ 2: ข้อกฎหมายที่ผ่านการแก้ไขเพิ่มเติม

ลำดับที่	ชื่อคุณลักษณะ	ตัวอย่าง	คำอธิบาย
1	lawid	CC-326-01	หมายเลขไอดีของข้อกฎหมาย มีที่มาจาก lawcode ต่อด้วย lawsection ต่อด้วยจำนวนครั้งในการแก้ไข กรณีนี้ คือ การแก้ไขครั้งที่ 1 หรือครั้งแรก
2	lawcode	CC	รหัสชื่อกฎหมาย ในที่นี้ CC มาจาก Criminal Code หมายความว่า ประมวลกฎหมายอาญา
3	lawsection	326	มาตราของข้อกฎหมาย แสดงว่าข้อกฎหมายนี้ คือ ประมวลกฎหมายอาญา มาตรา 326
4	category	2-11-3	หมวดหมู่ของมาตราในกฎหมายนั้น ในที่นี้ คือ ภาค 2 ลักษณะ 11 หมวด 3
5	effectivedate	2519-10-21	วันที่ข้อกฎหมายนี้มีผลบังคับใช้ นั่นคือ วันที่ 21 ตุลาคม พศ..2519
6	enddate	2535-02-25	วันที่ข้อกฎหมายนี้สิ้นสุดการบังคับใช้ นั่นคือ วันที่ 25 กุมภาพันธ์ พ.ศ.2535
7	editreference	คำสั่งของคณะปฏิรูปการปกครองแผ่นดิน ฉบับที่ 41 ลงวันที่ 21 ตุลาคม พ.ศ.2519	กฎหมายที่แก้ไขเพิ่มเติมมาตรา 326 เป็นครั้งแรก
8	content	ผู้ใดใส่ความผู้อื่นต่อบุคคลที่สามโดยประการที่น่าจะทำให้ผู้อื่นนั้นเสียชื่อเสียง ถูกดูหมิ่น หรือถูกเกลียดชัง ผู้นั้นกระทำความผิดฐานหมิ่นประมาท ต้องระวางโทษจำคุกไม่เกินหนึ่งปี หรือปรับไม่เกินสองพันบาท หรือทั้งจำทั้งปรับ	เนื้อหาข้อกฎหมาย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บรรณานุกรม

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียน

ชื่อ-สกุล	นายกานกวิญจน์ โค้วสีหวัฒน์
วัน เดือน ปี เกิด	18 กรกฎาคม 2535
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	วิศวกรรมศาสตรบัณฑิต เกียรตินิยมอันดับสอง (สาขาวิชาวิศวกรรมคอมพิวเตอร์) จุฬาลงกรณ์มหาวิทยาลัย พ.ศ.2558 นิติศาสตรบัณฑิต เกียรตินิยมอันดับสอง มหาวิทยาลัยรามคำแหง พ.ศ.2555 สอบไล่ได้เป็นเนติบัณฑิตจากสำนักอบรมศึกษากฎหมายแห่งเนติบัณฑิตยสภา พ.ศ.2558
ที่อยู่ปัจจุบัน	768/110 หมู่บ้านพฤกษาวิลล์ 73 ซอยพัฒนาการ 38 ถนนพัฒนาการ แขวงสวนหลวง เขตสวนหลวง กรุงเทพมหานคร 10250
ผลงานตีพิมพ์	K. Kowsrihawat, P. Vateekul and P. Boonkwan, "Predicting Judicial Decisions of Criminal Cases from Thai Supreme Court Using Bi-directional GRU with Attention Mechanism," in Proceedings of the 5th Asian Conference on Defense Technology, ACDT 2018, 2018.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University