การพัฒนาการค้นคืนข้อมูลข้ามมีเดียบนดัชนีแฮชโค้ด K-means

The development of cross-media retrieval on the K-means hash code index

ศราวุธ มากชิต

งานวิจัยนี้ได้รับทุนสนับสนุนการวิจัย
จากสำนักงานคณะกรรมการส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม (สกสว.)
(กันยายน 2565)

การพัฒนาการค้นคืนข้อมูลข้ามมีเดียบนดัชนีแฮชโค้ด K-means

The development of cross-media retrieval on the K-means hash code index

ดร.ศราวุธ มากชิต
สาขาคอมพิวเตอร์ คณะครุศาสตร์
มหาวิทยาราชภัฏสุราษฎร์ธานี

งานวิจัยนี้ได้รับทุนสนับสนุนการวิจัย
จากสำนักงานคณะกรรมการส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม (สกสว.)
(กันยายน 2565)

บทสรุปผู้บริหาร

ในปัจจุบันนี้ข้อมูลทางด้านมัลติมีเดียได้เพิ่มขึ้นอย่างรวดเร็วในแต่ละวัน เช่น ข้อความ (Texts) รูปภาพ (Images) สื่อวีดิโอ (Videos) เสียง (Audios) และข้อมูลสามมิติ (3Ds) เป็นต้น เนื่องด้วยความสะดวกในการใช้อุปกรณ์ต่าง ๆ เพื่อบันทึกข้อมูล อีกทั้งระบบเครือข่ายที่ได้มีการพัฒนาอย่างรวดเร็วเช่นกัน การสืบค้นข้อมูลทางด้านมัลติมีเดียหรือข้ามมีเดียจึงเป็นเรื่องสำคัญมากในยุคนี้ โดยที่เราสามารถค้นคืนข้อมูลจากการใช้คำค้นหรือสื่อแต่ละประเภทเพื่อสืบค้นข้อมูลที่มีเนื้อหาใกล้เคียงกันได้ (Information retrieval) และการค้นคืนข้อมูลข้ามมีเดียหรือข้ามประเภทสื่อ (Cross-media retrieval) ก็เป็นสิ่งที่น่าสนใจมากสำหรับผู้ใช้งานในทุกวันนี้ ซึ่งนักวิจัยและนักพัฒนาระบบก็ให้ความสนใจมากเช่นเดียวกัน

ในการค้นคืนข้อมูลด้วยข้อมูลดิบหรือข้อมูลจริง (Raw data or real data) และค้นหาแบบละเอียดถี่ถ้วน (Exhaustive search) ไม่สามารถทำได้ในโลกแห่งความเป็นจริง (Real world application) ดังนั้นการเข้ารหัสข้อมูล (Encoding) เพื่อให้ข้อมูลมีขนาดเล็กลง (Compact code) และการจัดการดัชนี (Index) ข้อมูลจึงเป็นสิ่งจำเป็นอย่างยิ่งที่จะต้องได้รับการพัฒนาปรับปรุงให้มีประสิทธิภาพมากขึ้น ในการเข้ารหัสข้อมูล (Embedding) จะทำให้ข้อมูลบางส่วนขาดหายไป (Information loss) ในขณะที่ทำการเข้ารหัส ซึ่งมีผลทำให้ผลการค้นคืนข้อมูลจะมีประสิทธิภาพลดลง แต่จะสามารถประหยัดเวลาในการสืบค้นข้อมูลได้หลายเท่าเช่นกัน ซึ่งถือว่าเป็นข้อดีในเรื่องของเวลา โดยเวลาที่ลดลงจะขึ้นอยู่กับชนิดและขนาดของข้อมูลหลังจากการเข้ารหัส การค้นหาข้อมูลแบบละเอียดก็ยังเป็นปัญหาวิกฤตในระบบสืบค้นของชุดข้อมูลขนาดใหญ่ (Large-scale dataset) ถึงแม้จะมีการเข้ารหัสข้อมูลแล้วก็ตาม

การจัดทำดัชนีในการสืบค้นเป็นเรื่องที่สำคัญไม่น้อยไปกว่าการเข้ารหัสข้อมูล ซึ่งงานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการเข้ารหัสข้อมูลด้วยเทคนิค K-means clustering ในใช้ในการสร้างตารางดัชนี และพัฒนาปรับปรุงการทำตารางดัชนีแบบกลับด้าน (Inverted index) ให้มีประสิทธิภาพเพิ่มขึ้น ด้วยวิธีการดำเนินงานวิจัย คือ ใช้วิธีการคำนวณค่าความสัมพันธ์ (Relevant score) หรือค่าความน่าจะเป็น (Probability) ของคู่ความสัมพันธ์ข้อมูลในการค้นหาข้อมูลของแต่ละคำค้น (Query) กับค่าคำตอบที่ถูกต้องของแต่ละกลุ่มข้อมูล (Cluster) ที่จัดกลุ่มด้วย K-means เพื่อสร้างรูปแบบ (Pattern) สำหรับสร้างโมเดลการทำนาย (Prediction model) โดยใช้คำค้นที่เป็นข้อมูลดิบจับคู่กับค่าความสัมพันธ์ของแต่ละกลุ่มข้อมูลโดยผ่านวิธีเรียนรู้เชิงลึก (deep learning) เพื่อจัดเรียงลำดับข้อมูลดัชนีใหม่ตามค่าความสัมพันธ์ โดยผลลัพธ์ของงานวิจัยนี้ คือ สามารถเพิ่มความถูกต้องในการค้นคืนข้อมูลและลดเวลาในการประมวลผลได้ และผู้ที่สนใจสามารถนำแนวคิดดังกล่าวไปประยุกต์ใช้งานในระบบที่ใช้งานจริงในการสืบค้นข้อมูลข้ามมีเดีย หรือสามารถนำแนวคิดไปทดลองประยุกต์ใช้งาน

กับเทคนิคการเข้ารหัส Hash code ด้วยวิธีการอื่น ๆ มาใช้ในการสร้างดัชนีและคำนวณค่าความสัมพันธ์และทดสอบความถูกต้อง และสามารถทดลองใช้กับจำนวนกลุ่มหรือจำนวนบิตที่สูงขึ้นในการสร้างตารางดัชนีและการเรียนรู้ โดยใช้เครื่องคอมพิวเตอร์ที่มีสมรรถนะที่สูงกว่าการทดลองครั้งนี้ และสามารถทดลองประยุกต์ใช้งานหลาย K-means ที่มากกว่า 2 รหัส เพื่อดูผลลัพธ์ในมุมมองที่แตกต่างกันออกไป

บทคัดย่อ

หัวข้อวิจัย การพัฒนาการค้นคืนข้อมูลข้ามมีเดียบนดัชนีแฮชโค้ด K-means

ผู้ดำเนินการวิจัย ดร.ศราวุธ มากชิต

หน่วยงาน สาขาคอมพิวเตอร์ศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสุราษฎร์ธานี

ปีการศึกษา 2565

เทคนิค Vector Quantization (VQ) ได้ถูกนำมาใช้กันอย่างแพร่หลายในการจัดการกับปัญหาและอุปสรรคของการจัดเก็บและเวลาในการประมวลผลข้อมูล สำหรับการค้นหาเพื่อนบ้านที่ใกล้ที่สุด (Nearest Neighbors: NN) ในแอปพลิเคชันต่าง ๆ ของข้อมูลมัลติมีเดีย ซึ่งอัลกอริทึม K-means clustering เป็นวิธีการหนึ่งของ VQ เพื่อจัดการกับปัญหาดังกล่าว ปัจจุบันการค้นคืนข้อมูลข้ามมีเดีย (Cross-media retrieval) มีความสำคัญมากขึ้น เนื่องจากการเติบโตอย่างรวดเร็วของข้อมูลทางด้านมัลติมีเดีย เช่น ข้อความ รูปภาพ วิดีโอ เสียง และ ข้อมูล 3 มิติ เราสามารถค้นคืนข้อมูลโดยการส่งคำถาม (query) ของสื่อประเภทหนึ่งเพื่อค้นคืนสื่ออีกประเภทหนึ่ง ถึงแม้ว่าจะมีการนำเสนอเทคนิค VQ หรืออัลกอริธึมการทำแฮชชิง (Hashing) จำนวนมาก เพื่อรหัสขนาดกะทัดรัด (Compact code) โดยมีวัตถุประสงค์ในการทำงานแบบเรียลไทม์ แต่ถ้าจะค้นหาอย่างละเอียดถี่ถ้วน (Exhaustive search) นั้นไม่สามารถทำได้ในโลกแห่งความเป็นจริง และการคำนวณระยะทางแฮมมิง (Hamming distance) ก็ให้ผลลัพธ์ที่คลาดเคลื่อนสูง เนื่องจากมีการสูญเสียข้อมูลขณะทำการเข้ารหัส (Quantization loss) ในงานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาตารางดัชนีแบบกลับด้านด้วยรหัส K-means และ 2) นำเสนอแนวทางและวิเคราะห์ประสิทธิภาพการค้นคืนข้อมูลข้ามมีเดียแบบใหม่ที่ผ่านการเรียนรู้เชิงลึก (DNN) โดยใช้วิธีการเรียนรู้คุณสมบัติของข้อมูลมัลติมีเดีย (Multimedia features) เพื่อการค้นคืนข้อมูลข้ามมีเดีย วิธีการที่นำเสนอเป็นวิธีการค้นหาแบบใหม่ที่ใช้ค่าความสัมพันธ์ (Relevant score) ของตารางดัชนีแบบกลับด้าน (Inverted index) ด้วยรหัสกลุ่ม K-means จับคู่กับคำถามดิบ (Raw query) โดยผ่านการเรียนรู้ด้วยโครงข่ายประสาทเทียมเชิงลึก (DNN) เพื่อปรับปรุงประสิทธิภาพของดัชนี K-means แบบดั้งเดิม การทดสอบได้ดำเนินการกับชุดข้อมูลมัลติมีเดียมาตรฐาน 4 ชุด ผลลัพธ์ที่ได้แสดงให้เห็นถึงประสิทธิผลที่ช่วยเพิ่มประสิทธิภาพเมื่อเทียบกับการค้นคืนแบบพื้นฐาน (baseline) ด้วยรหัส K-means ทั้งด้านความถูกต้องและเวลา ซึ่งสามารถกล่าวสรุปได้ว่า โมเดลการเรียนรู้ DNN สามารถเรียนรู้คุณสมบัติข้อมูลคำถามดิบของข้อมูล

มัลติมีเดียได้ดีในการเชื่อมโยงกับค่าคะแนนความสัมพันธ์ของแต่ละกลุ่ม K-means ได้อย่างมีประสิทธิภาพ

คำสำคัญ – การจัดทำดัชนีกลับด้าน การค้นคืนข้อมูลข้ามมีเดีย การค้นหาเพื่อนบ้านที่ใกล้ที่สุด ดัชนีรหัสไบนารี่

Abstract

Research Title The development of cross-media retrieval on the K-means hash code index

Researcher Sarawut Markchit, Ph.D.

Organization Computer Education, Faculty of Education, Suratthani Rajabhat University

Academic Year 2022

Vector quantization (VQ) has lately become commonly employed for Nearest Neighbors (NN) search in multimedia retrieval applications to overcome the limitation of the storage and calculation time. One of VQ technique called K-means clustering can be used to solve the problems. Due to the growing rise of multimedia data such as text view, photo view, video view, audio view, and 3D view, cross-media retrieval has become increasingly vital. We can submit a query of each media view to retrieve the results of another media view. Despite the fact that several VQ approaches or hashing algorithms have been developed to construct compact binary codes. An exhaustive search is unfeasible in real-time, and Hamming distance computation in the Hamming space produces erroneous results because of quantization loss.

The objectives of this work are, 1) to develop inverted index tables with K-means code; and 2) to propose the new way and analyze the effective of cross-media retrieval by applying deep learning. The method had learned each multimedia features and search for another kinds of media view in the K-means index. As a result, we offer a unique search strategy for K-means hash codes that employs a relevant score of index structure. We create an inverted index table based on K-means clusters and train a neural network by matching a raw query with the relevant score of each cluster to increase the accuracy and efficiency of the original K-means index. Experiments are conducted on four benchmark multimedia datasets. Our findings suggest that the efficacy improves baseline performance in accuracy and computation time. We can conclude that The DNN learning model is able to learn

the raw question data properties of multimedia data effectively in association with the correlation score values of each K-means group.

Keywords - inverted indexing, Cross-media retrieval, nearest neighbor search, binary code index

กิตติกรรมประกาศ

งานวิจัยเรื่องการพัฒนาการค้นคืนข้อมูลข้ามมีเดียบนดัชนีแฮชโค้ด K-means คงสำเร็จลุล่วงไม่ได้ ถ้าไม่ได้รับความช่วยเหลือสนับสนุนจากหน่วยงานและบุคคลจำนวนหนึ่ง ซึ่งผู้วิจัยขอกล่าวขอบคุณดังนี้

ขอขอบคุณครอบครัวที่ให้กำลังใจนักวิจัยให้มีความอดทน ขยัน ต่อสู้กับปัญหา และคอยสนับสนุนตลอดมา ขอขอบคุณเพื่อนร่วมงานและมหาวิทยาลัยราชภัฏสุราษฎร์ธานี ที่สนับสนุนเรื่องเวลางาน เครื่องคอมพิวเตอร์ และวัสดุอุปกรณ์ต่าง ๆ ในการดำเนินการวิจัย

สุดท้ายผู้วิจัยขอขอบคุณทุนอุดหนุนการวิจัยจากสำนักงานคณะกรรมการส่งเสริมวิทยาศาสตร์ วิจัย และนวัตกรรม (สกสว.) ประเภททุนสนับสนุนงานมูลฐาน (Fundamental Fund: FF) ปีงบประมาณ 2565 ในการทำวิจัยครั้งนี้

ศราวุธ มากชิต
กันยายน 2565

สารบัญ

	หน้า
บทสรุปผู้บริหาร	ค
บทคัดย่อภาษาไทย	จ
บทคัดย่อภาษาอังกฤษ	ช
กิตติกรรมประกาศ	ฌ
สารบัญ	ญ
สารบัญตาราง	ฏ
สารบัญภาพ	ฑ
บทที่ 1 บทนำ	
ความเป็นมาและความสำคัญ	1
วัตถุประสงค์	2
กรอบแนวคิด	3
สมมติฐานการวิจัย	3
ขอบเขตการวิจัย	3
ข้อจำกัด	4
นิยามศัพท์	4
ประโยชน์ที่ว่าจะได้รับ	5
บทที่ 2 แนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง/วรรณกรรมที่เกี่ยวข้อง	
แนวคิด ทฤษฎีที่เกี่ยวข้อง	6
งานวิจัยที่เกี่ยวข้องกับปัญหาการวิจัย	11
บทที่ 3 วิธีดำเนินการวิจัย	
รูปแบบการวิจัย	16
ประชากรและกลุ่มตัวอย่าง	16
เครื่องมือที่ใช้ในการรวบรวมข้อมูล	19
วิธีการเก็บรวบรวมข้อมูล	20
สถิติที่ใช้ในการวิเคราะห์ข้อมูล	21
ขั้นตอนการทดลอง	23

บทที่ 4 ผลการวิจัย

ผลการค้นคืนข้อมูลด้วยตารางดัชนีแบบกลับด้านของรหัส K-means 29

ประสิทธิภาพการค้นคืนข้อมูลข้ามมีเดีย 31

บทที่ 5 สรุป อภิปรายผล และข้อเสนอแนะ 41

เอกสารอ้างอิง/บรรณานุกรม 43

ภาคผนวก 47

ภาคผนวก ก ข้อมูลและเอกสารเกี่ยวข้องกับงานวิจัย 48

ภาคผนวก ข ประวัติผู้วิจัย 54

สารบัญตาราง

ตารางที่ **หน้า**

3.1 รูปแบบการเปรียบเทียบศึกษาการวิจัย 16

3.2 รายละเอียดชุดข้อมูล WIKI MIRFlickr และ NUS-WIDE 17

3.3 รายละเอียดชุดข้อมูล XMedia 17

3.4 สรุปรายการชุดข้อมูลมัลติมีเดียที่นิยมใช้กันอย่างแพร่หลาย 20

4.1 เปรียบเทียบค่า MAP ของผลการค้นคืนข้อมูลด้วยรหัส K-means ชุดข้อมูล Wiki 29

4.2 เปรียบเทียบค่า MAP ของผลการค้นคืนข้อมูลด้วยรหัส K-means ชุดข้อมูล NUS-WIDE 30

4.3 การเปรียบเทียบค่า precision ของ top 1 ถึง 100 อันดับแรกของผลการค้นคืนข้อความโดยใช้คำค้นด้วยรูปภาพ (I -> T) ด้วยรหัสไบนารี่ 8 และ 10 บิต สำหรับชุดข้อมูล Wiki 31

4.4 การเปรียบเทียบค่า recall ของ top 1 ถึง 100 อันดับแรกของผลการค้นคืนข้อความโดยใช้คำค้นด้วยรูปภาพ (I -> T) ด้วยรหัสไบนารี่ 8 และ 10 บิต สำหรับชุดข้อมูล Wiki 31

4.5 การเปรียบเทียบค่า precision ของ top 1 ถึง 100 อันดับแรกของผลการค้นคืนรูปโดยใช้คำค้นด้วยข้อความ (T -> I) ด้วยรหัสไบนารี่ 8 และ 10 บิต สำหรับชุดข้อมูล Wiki 32

4.6 การเปรียบเทียบค่า recall ของ top 1 ถึง 100 อันดับแรกของผลการค้นคืนรูปโดยใช้คำค้นด้วยข้อความ (T -> I) ด้วยรหัสไบนารี่ 8 และ 10 บิต สำหรับชุดข้อมูล Wiki 32

4.7 การเปรียบเทียบค่า precision และ recall ของ top 1 ถึง 100 อันดับแรกของผลการค้นคืนรูปโดยใช้คำค้นด้วยข้อความ (I -> T) ด้วยรหัสไบนารี่ 8 บิต สำหรับชุดข้อมูล MIRFlickr 33

4.8 การเปรียบเทียบค่า precision และ recall ของ top 1 ถึง 100 อันดับแรกของผลการค้นคืนรูปโดยใช้คำค้นด้วยข้อความ (T -> I) ด้วยรหัสไบนารี่ 8 บิต สำหรับชุดข้อมูล MIRFlickr 33

4.9 การเปรียบเทียบค่า MAP ของ top 1 ถึง 50 อันดับแรกของผลการค้นคืนข้อความโดยใช้คำค้นด้วยรูปภาพ (I -> T) ด้วยรหัสไบนารี่ 5 8 และ 10 บิต สำหรับชุดข้อมูล Wiki 34

4.10 การเปรียบเทียบค่า MAP ของ top 1 ถึง 50 อันดับแรกของผลการค้นคืนรูปโดยใช้คำค้นด้วยข้อความ (T -> I) ด้วยรหัสไบนารี่ 8 และ 10 บิต สำหรับชุดข้อมูล Wiki 34

4.11 การเปรียบเทียบค่า MAP ของ top 1 ถึง 50 อันดับแรกของผลการค้นคืนข้ามมีเดียด้วยรหัสไบนารี่ 8 บิต สำหรับชุดข้อมูล Xmedia 35

4.12 การเปรียบเทียบค่า MAP ของ top 1 ถึง 50 อันดับแรกของผลการค้นคืนข้ามมีเดียด้วยรหัสไบนารี่ 8 บิต สำหรับชุดข้อมูล MIRFlickr 36

4.13 การเปรียบเทียบค่า MAP ของ top 1 ถึง 50 อันดับแรกของผลการค้นคืนข้ามมีเดียด้วยรหัสไบนารี่ 8 บิต สำหรับชุดข้อมูล NUS-WIDE 37
4.14 การเปรียบเทียบค่า MAP ของ top 1 ถึง 50 อันดับแรกของผลการค้นคืนข้ามมีเดียด้วยรหัสไบนารี่ 9 บิต สำหรับชุดข้อมูล NUS-WIDE 38
4.15 แสดงค่า MAP ของ 1 รายการแรกของผลการค้นคืนแบบ Multi-K-means 5 บิต บนชุดข้อมูล Wiki Xmedia และ MIRFlickr 39
4.16 การเปรียบเทียบเวลาที่ใช้ในการค้นคืนข้อมูล 10 บิต หน่วยเป็นวินาที ของผลลัพธ์ 50 อันดับแรก 39

สารบัญภาพ

ภาพที่ **หน้า**

1.1 กรอบแนวคิดการวิจัย 3
2.1 แสดงการจัดกลุ่มข้อมูลด้วย K-means 7
3.1 การจัดกลุ่มและสร้างดัชนีข้อมูล 21
3.2 แสดงกรอบ (framework) ของการค้นคืนข้อมูลของกระบวนการวิธีที่นำเสนอ 23
3.3 คะแนนความสัมพันธ์ 24
3.4 โมเดลการทำนาย 24
3.5 ขั้นตอนการค้นคืนข้อมูล 25
3.6 การประยุกต์ใช้หลายดัชนี K-means (Multi-K-means) 25

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญ

การค้นหาเพื่อนบ้านที่ใกล้ที่สุด เป็นพื้นฐานในการเรียนรู้ของเครื่อง (Machine learning: ML) และการสืบค้นข้อมูลสารสนเทศ (Information retrieval: IR) ซึ่งในปัจจุบันการประยุกต์ใช้การค้นหาเพื่อนบ้านที่ใกล้ที่สุดในการสืบค้นข้อมูลมัลติมีเดีย (Multimedia retrieval) ได้รับความสนใจในการพัฒนางานวิจัยและการประยุกต์ใช้งานเป็นอย่างมาก การสืบค้นข้อมูลมัลติมีเดียมีความสำคัญมากขึ้นเรื่อย ๆ เนื่องจากผู้ใช้สามารถดึงผลลัพธ์ด้วยสื่อประเภทต่าง ๆ ได้ โดยปกติข้อมูลมัลติมีเดียมีหลายรูปแบบ ซึ่งรูปแบบเหล่านั้นอาจให้ข้อมูลเชิงความหมาย (Semantic information) ที่สัมพันธ์กัน เช่น คู่แท็กวิดีโอ (Video-tag pairs) ใน YouTube และคู่ข้อความกับรูปภาพ (Image-text pairs) ในเว็บไซต์ Flickr [1]

ในช่วงไม่กี่ปีที่ผ่านมาการทำ Vector Quantization: VQ และ เทคนิคแฮชชิ่ง (Hashing technique) ได้รับความนิยมอย่างมากในการพัฒนางานวิจัย เนื่องจากมีประสิทธิภาพในการค้นคืนและจัดทำดัชนีข้อมูลมัลติมีเดียที่มีจำนวนหลายมิติ (High-dimensional) และขนาดใหญ่ (Large-scale) เนื่องด้วยการเพิ่มขึ้นอย่างรวดเร็วของข้อมูลมัลติมีเดียจึงเป็นไปไม่ได้ที่จะใช้วิธีการค้นหาข้อมูลอย่างละเอียด (Exhaustive search) ซึ่งวิธีการนี้จะใช้ทรัพยากรในการคำนวณมากมหาศาลสำหรับชุดข้อมูลขนาดใหญ่ (Large-scale dataset) เพื่อแก้ไขปัญหานี้ การสืบค้นข้อมูลมัลติมีเดียส่วนใหญ่จะใช้เทคนิคการแฮชชิ่ง หรือ เทคนิค VQ เพื่อสร้างข้อมูลที่กะทัดรัดในการแทนข้อมูลเดิม (Compact data representations) โดยเป้าหมายของการทำคือการเข้ารหัสข้อมูลจากดั้งเดิม (Original space) เป็นรหัสไบนารี (Binary hash codes) ใน Hamming space โดยทั่วไปแล้วจะใช้ประโยชน์จากความสัมพันธ์ของข้อมูล และการกระจายข้อมูลพื้นฐานโดยใช้ฟังก์ชันแฮช (Hash functions) เพื่อสร้างรหัสไบนารีที่คล้ายกันสำหรับข้อมูลที่คล้ายกัน ทำให้สามารถค้นหาข้อมูลเพื่อนบ้านที่ใกล้ที่สุดด้วยการคำนวณระยะห่างระหว่างรหัสไบนารี (Hamming distance) ได้อย่างรวดเร็วผ่านการดำเนินการระดับบิตที่ฮาร์ดแวร์รองรับ (Hardware-supported bit operations) โดยใช้หน่วยความจำน้อยที่สุด อย่างไรก็ตามการใช้เทคนิคการแฮชชิ่งก็ต้องเผชิญกับปัญหาวิกฤต (Critical problem) คือการสูญเสียรายละเอียดข้อมูล (Quantization loss) หลังจากการเข้ารหัสข้อมูล แม้ว่าจะมีการนำเสนอกลยุทธ์ต่าง ๆ ของอัลกอริธึม (Algorithms) การแฮชชิ่งในการเรียนรู้ต่าง ๆ เพื่อลดการสูญเสียข้อมูล แต่ก็ยังมีช่องว่างในข้อมูลขนาดใหญ่ (Large information gap) ที่หลีกเลี่ยงไม่ได้ระหว่างเวกเตอร์มูลค่าจริงกับรหัสไบนารีที่เกี่ยวข้อง การค้นหาเพื่อนบ้านที่ใกล้ที่สุดของรหัสไบนารีใน

Hamming space จึงมีความแม่นยำน้อยกว่าของข้อมูลจริง (Real-valued) ในการค้นหาข้อมูลด้วย Euclidean space [2] [3]

ปัจจุบันนี้ได้มีการนำเสนอวิธีการจัดทำดัชนีหลายวิธีสำหรับชุดข้อมูลที่มีหลายมิติและการค้นคืนข้อมูลข้ามมีเดีย แต่ในการพัฒนาอัลกอริธึมและโค้ดจะมีความซับซ้อนและยุ่งยาก บางครั้งความซับซ้อนนี้ก็ไม่สามารถเพิ่มประสิทธิภาพการจับคู่ได้ [1] ด้วยเหตุนี้งานวิจัยนี้จึงมีแนวคิดในการนำเสนอวิธีการใหม่ โดยย้อนกลับไปดูวิธีการที่ง่ายกว่าในการจัดทำดัชนีข้อมูลมัลติมีเดียและเสนอการจัดทำดัชนีที่เรียบง่าย แต่มีประสิทธิภาพสำหรับชุดข้อมูลที่มีมิติสูง โดยใช้ตารางดัชนีกลับด้านแบบใหม่จากการใช้ K-means ในการจัดกลุ่ม (Cluster) ข้อมูล เพื่อสร้างตารางดัชนี (Index code) โดยตารางดัชนีที่นำเสนอนี้สามารถใช้ประโยชน์จากจำนวน *K* (จำนวนกลุ่ม) ที่น้อยมาเป็นรหัสดัชนี โครงสร้างดัชนีถูกสร้างขึ้นโดยการจัดกลุ่มข้อมูลทั้งหมดที่มีกลุ่มหรือรหัสดัชนีเดียวกันไว้ในรายการของตารางแบบกลับด้าน (Inverted table) [5] เดียวกัน ผลความถูกต้องของการค้นคืนข้อมูลด้วยตารางดัชนีของ K-means ยังมีประสิทธิภาพน้อย ดังนั้นในส่วนของการค้นคืนข้อมูลของแนวคิดที่นำเสนอจะมีการพิจารณาจากคำค้นข้อมูลดิบหรือดั้งเดิม (Raw query หรือ original query) ของสื่อประเภทหนึ่งโดยการประมาณค่าความเกี่ยวข้องของรหัสดัชนีของสื่ออีกประเภทหนึ่งของแต่ละรหัสจากค่าความน่าจะเป็นของคำเฉลยเพื่อนบ้านที่ใกล้ที่สุด (Ground truth) การประมาณค่าจะได้จากแบบจำลองการทำนาย (Prediction model) ที่เรียนรู้แบบไม่เชิงเส้น (Nonlinear mapping) ของการเรียนรู้เชิงลึก (Deep learning) ระหว่างคำค้นข้อมูลดิบของข้อมูลแต่ประเภทข้อมูลกับตารางดัชนี (Index space) ของสื่ออีกประเภทหนึ่ง จากนั้นจะสำรวจตารางดัชนี (หรือกลุ่ม Cluster) จากรหัสดัชนีที่มีค่าทำนายสูงสุด (Top rank index codes) เพื่อดึงข้อมูลที่จะนำมาใช้เป็นตัวแทนหรือคู่แข่ง (Candidates) ที่มีคุณภาพสูง (ข้อมูลถูกต้อง) การประเมินประสิทธิภาพตารางดัชนีใหม่ที่นำเสนอจะใช้ชุดข้อมูลมัลติมีเดียมาตรฐาน (Benchmark datasets) ที่ใช้กันอย่างแพร่หลายโดยการประยุกต์ใช้ K-means ในการสร้างจัดกลุ่มข้อมูล

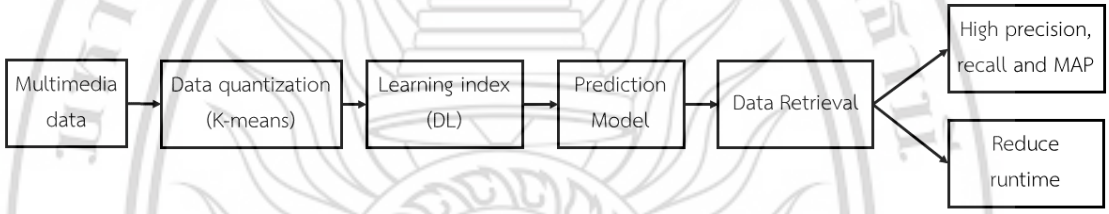
1.2 วัตถุประสงค์

1) เพื่อสร้างตารางดัชนีแบบกลับด้านด้วยรหัส K-means สำหรับชุดข้อมูลมัลติมีเดีย

2) เพื่อนำเสนอแนวทางและวิเคราะห์ประสิทธิภาพการค้นคืนข้อมูลข้ามมีเดียของตารางดัชนี K-means แบบเดิมด้วยตารางดัชนีกลับด้านแบบใหม่ที่ผ่านการเรียนรู้เชิงลึก (Deep learning)

1.3 กรอบแนวคิด

จากภาพที่ 1.1 แสดงกรอบแนวคิดของงานวิจัย โดยงานวิจัยนี้พัฒนาโมเดล (Prediction model) จากการประยุกต์ใช้งานเรียนรู้เชิงลึก (Deep learning) โดยเรียนรู้จากข้อมูลมัลติมีเดียตัวอย่าง (Multimedia training data) จากการจัดกลุ่มหรือเข้ารหัสข้อมูล (Data quantization) ด้วยเทคนิค K-means เพื่อใช้ในการสร้างตารางดัชนี (Index table) หลังจากได้โมเดลการเรียนรู้แล้วสามารถนำโมเดลมาประยุกต์ใช้ในการทำนายดัชนีหรือกลุ่มข้อมูลที่เป็นคำตอบ (Data retrieval) ของคำถามใหม่ (Query) ที่ต้องการ สามารถวัดประสิทธิภาพการค้นคืนข้อมูลได้ 2 แบบ คือ 1) ค่าความถูกต้อง (Precision, Recall and MAP) และ 2) เวลาในการสืบค้น (Time complexity)



ภาพที่ 1.1 กรอบแนวคิดการวิจัย

1.4 สมมติฐานการวิจัย

การเรียนรู้เชิงลึกเพื่อหาความสัมพันธ์ระหว่างคำค้นข้อมูลดิบกับค่าความสัมพันธ์ของแต่ละกลุ่มข้อมูล หรือตารางดัชนี สามารถเพิ่มประสิทธิภาพการค้นคืนข้อมูลแบบข้ามมีเดียผ่านตารางดัชนีกลับด้านได้

1.5 ขอบเขตการวิจัย

ประชากร ชุดข้อมูลมัลติมีเดียมาตรฐานที่เป็นที่นิยม 4 ชุดข้อมูลที่ใช้ในการดำเนินการทดลอง ได้แก่ ชุดข้อมูล Wiki [6] ชุดข้อมูล MIRFlickr [7] ชุดข้อมูล NUS-WIDE [8] และ ชุดข้อมูล XMedia [9]

กลุ่มตัวอย่าง ชุดข้อมูลทุกชุดข้อมูลจะถูกสุ่มตัวอย่างเพื่อแบ่งออกเป็นชุดข้อมูลที่ใช้ในการเรียนรู้ของโมเดล (Training set) และที่ใช้ในการทดสอบโมเดล (Testing set) โดยมีรายละเอียดการแบ่งข้อมูลแสดงไว้ในบทที่ 3

ข้อจำกัด

ปัจจุบันมีชุดข้อมูลมัลติมีเดียที่มีหลากหลายประเภท (มากกว่า 2 ประเภท ข้อความและรูปภาพ) ที่มีขนาดใหญ่มีให้ทดสอบน้อย เครื่องคอมพิวเตอร์ที่นักวิจัยใช้ในการทดลองมีประสิทธิภาพไม่สูง จึงเป็นข้อจำกัดในการทำวิจัยครั้งนี้

นิยามศัพท์

การค้นคืนข้อมูลข้ามมีเดีย (Cross-media retrieval) คือ การใช้คำค้น (Query) ด้วยสื่อมีเดียประเภทหนึ่งเพื่อสืบค้นข้อมูลในสื่ออีกประเภทหนึ่งที่มีความสัมพันธ์กัน

ดัชนีแฮชโค้ด (Hash code index) คือ รหัสที่นำมาใช้เป็นรหัสเพื่อสร้างตารางดัชนี

ค่าค่าความสัมพันธ์หรือความน่าจะเป็น (Relevant score หรือ Probability) คือ ค่าความสัมพันธ์หรือค่าความน่าจะเป็นของข้อมูลคำตอบหรือเพื่อนบ้านที่ใกล้เคียงกับคำถามในแต่ละกลุ่มข้อมูล (Cluster) โดยคำนวณจากจำนวนข้อมูลที่ถูกต้องหรือข้อมูลที่มีความสัมพันธ์กับคำถามหารด้วยจำนวนข้อมูลทั้งหมดในกลุ่มข้อมูลนั้น ๆ ผลรวมจะมีค่าเท่ากับหนึ่ง

เพื่อนบ้านที่ใกล้ที่สุด (Nearest neighbors: NN) คือ ข้อมูลคำตอบที่ใกล้เคียงหรือสัมพันธ์กับคำถามมากที่สุด จำนวนข้อมูลที่ต้องการของคำตอบขึ้นอยู่กับจำนวน k ที่กำหนดไว้

โมเดลการเรียนรู้หรือโมเดลการทำนาย (Learned or prediction model) คือ โมเดลที่ผ่านการเรียนรู้จากข้อมูลที่ใช้สอน (Training data) ข้อมูลที่ใช้สอนประกอบด้วยข้อมูลคำถามดิบ (Raw query) และคำตอบหรือกลุ่มข้อมูล (Answer or cluster) ทั้งสองข้อมูลจะถูกจับคู่ระหว่างคำถามและคำตอบเพื่อสอนให้โมเดลสามารถเรียนรู้และจดจำคุณลักษณะ (Feature) ของข้อมูลคำถามดิบ หลังจากเรียนรู้แล้วสามารถใช้โมเดลในการทำนายข้อมูลใหม่หรือคำถามใหม่ที่ยังไม่ได้ผ่านการเรียนรู้มาก่อน เพื่อใช้ในการทำนายคำตอบ (กลุ่มข้อมูลหรือตารางดัชนี) ที่ใกล้เคียงที่สุด

การค้นหาแบบละเอียด (Exhaustive search) คือ การค้นหาแบบทั้งหมดโดยที่คำถามแต่ละคำถามจะถูกเปรียบเทียบกับข้อมูลทั้งหมดในฐานข้อมูล (Reference set) เพื่อหาค่าความคล้ายคลึงหรือความเหมือน (Similarity) แล้วแสดงผลลัพธ์คำตอบที่มีความคล้ายกับคำถามมากที่สุดไปยังน้อยที่สุดตามลำดับ

ระยะทางดั้งเดิมหรือข้อมูลดิบ (Original distance) คือ ระยะทางหรือความแตกต่างระหว่าง 2 ข้อมูลที่มีความยาวเท่ากัน โดยที่จำนวนตำแหน่งที่มีสัญลักษณ์หรืออักขระที่แตกต่างกัน

คือจำนวนตัวอักษรที่คลาดเคลื่อนจากข้อมูลหนึ่งกับอีกข้อมูลหนึ่งบนชุดข้อมูลดิบ ค่าที่น้อยที่สุดคือชุดข้อมูลที่ความคล้ายกันมากที่สุด เช่น การหาค่า Euclidean distance

ระยะทางแฮมมิง (Hamming distance) คือ ระยะทางหรือความแตกต่างระหว่างจำนวนบิตของเลขฐานสองหรือรหัสไบนารี่ที่มีความยาวเท่ากัน (XOR) ค่าความแตกต่างที่น้อยที่สุดแสดงถึงความคล้ายคลึงกันระหว่างเลขไบนารี่ทั้งสองชุดมากที่สุด

หลาย K-means (Multi-K-means) เป็นการนำรหัส K-means หลายรหัสมาเรียนรู้ผ่านโมเดลพร้อม ๆ กัน แล้วเอาผลลัพธ์ของแต่ละชุดมาโหวตเพื่อหาข้อมูลที่คล้ายคลึงกันมากที่สุด

ประโยชน์ที่คาดว่าจะได้รับ

หลังจากเสร็จสิ้นงานวิจัยนี้คาดว่าผลการทดลองที่จะได้รับ คือ วิธีการที่เสนอสามารถปรับปรุงประสิทธิภาพการค้นคืนข้อมูลข้ามมีเดียได้อย่างมีประสิทธิภาพทั้งในแง่ของความแม่นยำของผลลัพธ์และเวลาในการคำนวณ โดยตารางดัชนีกลับด้านที่เสนอจะมีประโยชน์ดังต่อไปนี้

1. โครงสร้างดัชนีที่สร้างขึ้นจะมีกระบวนการดึงข้อมูลที่สามารถจัดการเรื่องความซับซ้อนของเวลาเชิงเส้นย่อย (Sub-linear time complexity) ผ่านการค้นหาตารางแบบกลับด้าน เมื่อเทียบกับการค้นหาแบบละเอียด (Exhaustive search) ที่ใช้ความซับซ้อนของเวลาเชิงเส้น

2. เมื่อพิจารณาจากคำค้น รูปแบบการทำนายที่ผ่านการเรียนรู้ (Prediction model) ที่ใช้เพื่อประมาณค่าความเกี่ยวข้อง (Relevance scores) ของรหัสดัชนีคาดว่าน่าจะมีความแม่นยำสูงกว่าการคำนวณระยะทางแบบดั้งเดิมของแฮมมิง (Hamming distances) ที่สร้างจาก K-means

3. นักพัฒนาหรือนักวิจัยทั่วไปที่มีความสนใจสามารถนำแนวคิดไปพัฒนา ปรับปรุง หรือต่อยอดงานทางด้านการค้นคืนข้อมูลข้ามมีเดีย หรือประยุกต์ใช้งานศาสตร์อื่น ๆ ได้

บทที่ 2
แนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง/วรรณกรรมที่เกี่ยวข้อง

แนวคิด ทฤษฎี เอกสารที่เกี่ยวข้องกับงานวิจัยนี้สามารถแบ่งออกเป็น 3 หัวข้อดังต่อไปนี้

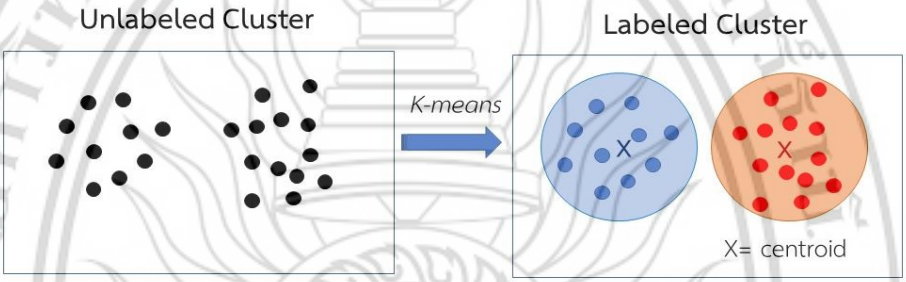
2.1 เทคนิคการเข้ารหัสข้อมูล (Data quantization)

ปัจจุบันนี้มีข้อมูลปริมาณที่มากและเพิ่มขึ้นทุกวัน อีกทั้งข้อมูลมีความหลากหลายรูปแบบหรืออาจจะอยู่ในรูปแบบที่รวมกันซึ่งเรียกว่าข้อมูลมัลติมีเดีย โดยความหมายของ “มัลติมีเดีย” มาจากคำว่า มัลติ (Multi) หมายถึง หลาย ๆ อย่างผสมรวมกัน และมีเดีย (Media) คือ สื่อ ข่าวสาร ช่องทางการสื่อสารเพื่อนำมารวมกัน เป็นคำว่า “มัลติมีเดีย” (Multimedia) ซึ่งเมื่อรวมกันแล้วมีความหมาย คือ การนำองค์ประกอบของสื่อชนิดต่าง ๆ มาผสมผสานรวมกัน ซึ่งประกอบด้วย ตัวอักษร (Text) ภาพนิ่ง (Image) ภาพเคลื่อนไหว (Animation) เสียง (Sound) และวีดีโอ (Video) โดยผ่านกระบวนการทางระบบคอมพิวเตอร์ เพื่อสื่อความหมายกับผู้ใช้อย่างมีปฏิสัมพันธ์ (Interactive Multimedia) และได้บรรลุตามวัตถุประสงค์การใช้งาน เนื่องด้วยทั้งขนาดและปริมาณข้อมูลจึงส่งผลต่อการจัดเก็บและการค้นคืนข้อมูล ซึ่งทำให้เทคนิคการเข้ารหัสและการจัดทำดัชนีมีความจำเป็นอย่างยิ่งในการจัดการกับข้อมูลทางด้านมัลติมีเดีย [1]

Data quantization เป็นเทคนิคการเข้ารหัสเพื่อแทนที่ข้อมูล (Data represented) ต้นฉบับ (Original data) ด้วยรหัสขนาดกะทัดรัด (Compact codes) ที่ใช้กันอย่างแพร่หลายในการค้นหาเพื่อนบ้านที่ใกล้ที่สุด (NN) อย่างมีประสิทธิภาพ สำหรับการค้นคืนข้อมูลมัลติมีเดียในมิติสูงและข้อมูลขนาดใหญ่ ทฤษฎีหรืองานวิจัยที่เกี่ยวกับการสร้างรหัสกะทัดรัดโดยทั่วไปจะแบ่งออกเป็น 2 ประเภท คือ การเรียนรู้แฮชชิง (Hash learning) และการหาปริมาณเวกเตอร์ (Vector quantization: VQ) ซึ่งความท้าทายหลักของ 2 วิธีนี้คือมีประสิทธิในการค้นคืนน้อยเนื่องจากเกิด Quantization error ในขณะที่ทำการเข้ารหัสข้อมูล [10]

ในแนวทางการเรียนรู้แฮชชิงชุดของฟังก์ชันแฮชจะเปลี่ยนข้อมูลต้นฉบับเป็นรหัสไบนารี่ การคำนวณระยะทางแฮมมิง (Hamming distance) ระหว่างรหัสไบนารี่สองรหัส สามารถประมวลผลได้เร็วมากเมื่อเทียบกับการคำนวณระยะทางแบบ Euclidian distance ระหว่างเวกเตอร์ข้อมูลจริง (real-valued vectors) สองชุดข้อมูลในข้อมูลต้นฉบับ (original space) ในส่วนของเทคนิค VQ จะมีการสูญเสียข้อมูลน้อยกว่าวิธีการเรียนรู้แฮชชิง แต่การเข้ารหัสและการคำนวณระยะทางจะประมวลผลได้ช้ากว่า หลักการสำคัญในการทำ VQ คือการสุ่มเวกเตอร์ที่จะใช้ในการเข้ารหัสไบนารี่

โดยที่เวกเตอร์อินพุต (input vectors) แต่ละตัวสามารถมีจำนวน *n* มิติได้ จะมีการแบ่งพาร์ติชันของเวกเตอร์ควอนไทเซอร์ (vector quantizer) ตามจำนวน *n* มิติที่ไม่ซ้ำกัน เวกเตอร์ถูกเข้ารหัสโดยการเปรียบเทียบกับโค้ดบุ๊ค (codebook) ที่ประกอบด้วยชุดของเวกเตอร์อ้างอิง (reference vectors) ที่จัดเก็บไว้ซึ่งเรียกว่า Codevectors การออกแบบ codebook เพื่อใช้แทนที่ (represents) ชุดเวกเตอร์อินพุตให้ได้ดีที่สุดจะเจอปัญหา NP-hard หมายความว่าจะต้องมีการค้นหาอย่างละเอียด (exhaustive search) เพื่อหารหัสคำศัพท์ที่ดีที่สุด และจะต้องเพิ่มจำนวน codewords เป็นทวีคูณถึงจะได้ผลคำตอบที่ดี [2]



ภาพที่ 2.1 แสดงการจัดกลุ่มข้อมูลด้วย K-means

ที่มา: ปรับปรุงจาก https://towardsdatascience.com/ [11]

จากภาพตัวอย่างที่ 2.1 เทคนิค K-means clustering [4] เป็นเทคนิคหนึ่งของการทำ VQ คือการเข้ารหัสข้อมูลวิธีหนึ่งโดยการจับกลุ่มข้อมูลที่คล้ายกันให้อยู่ในกลุ่มเดียวกัน (cluster) และจัดกลุ่มข้อมูลที่แตกต่างกันให้ห่างกันหรืออยู่คนละกลุ่มกัน ซึ่ง K-means เป็นวิธีการที่ง่ายและเป็นที่นิยมในการเรียนรู้แบบ unsupervised ของการเรียนรู้ของเครื่อง (machine learning: ML) ซึ่งโดยปกติ unsupervised อัลกอริทึมจะทำการอนุมาน (inferences) จากชุดข้อมูลโดยจะใช้เวกเตอร์อินพุตเท่านั้น (จุดสีดำ) โดยไม่มีติดป้ายกำกับหรือข้อมูลคำตอบ (labeled) ที่ใช้ในการเรียนรู้ อัลกอริทึม K-means clustering เป็นวิธีการจัดกลุ่มข้อมูลลงในเซลล์ Voronoi (ตัวอย่างคือกลุ่มสีฟ้าและสีแดง) ซึ่งจุดข้อมูลแต่ละจุดอยู่ในกลุ่มข้อมูล (cluster) ที่มีค่ากลาง centroid (X) ที่ใกล้ที่สุดและข้อมูลในกลุ่มนั้นจะถูกเข้ารหัสเป็นค่ากลาง หลักการทำงานของ K-means อัลกอริทึม (algorithm) จะเริ่มต้นด้วยการกำหนด (assigns) ข้อมูลแต่ละจุดให้กับกลุ่มข้อมูล แล้วจึงคำนวณค่าผลรวมที่น้อยที่สุดของ squared distance ระหว่างข้อมูลทั้งหมดในกลุ่ม เพื่อให้ข้อมูลที่มีลักษณะคล้ายกันอยู่ในกลุ่มเดียวกันด้วยสูตรการคำนวณดังต่อไปนี้

$$K(V) = \sum_{i=1}^{c} \sum_{j=1}^{ci} (||Xi - Vj||)^2, \quad (2.1)$$

โดยที่ ||Xi – Vj|| คือการคำนวณ Euclidean distance ระหว่าง x*i* และ v*j*

ci คือ จำนวนของจุดข้อมูลตัวที่ i*th* ของ cluster

c คือ จำนวนของ cluster หรือ จำนวนของ centroids

ขั้นตอนการทำงานของ K-mean algorithm มี 4 ขั้นตอนดังนี้

1. กำหนดจำนวนกลุ่ม (*K*) ตามที่ต้องการ ซึ่งค่านี้จะขึ้นอยู่กับความเหมาะสมของข้อมูล เช่น ถ้ากำหนด *n* กลุ่มหรือหมายถึงค่า K=*n* (กำหนดเป็น C*1*, C*2*, C*3*, ... C*n*) และสุ่มตำแหน่งแกน x,y ให้กับ C แต่ละตัว จะได้ว่า *C1*(x*1*,y*1*), C*2*(x*2*,y*2*), C*3*(x*3*,y*3*), ... C*n*(x*n*,y*n*)
2. ดูตำแหน่งของสมาชิกแต่ละจุดว่าอยู่ใกล้ค่า C ตัวไหนมากกว่าก็ให้เป็นสมาชิกของ C นั้น
3. ปรับ x,y ของ C แต่ละตัวใหม่ให้อยู่ตรงกลางของกลุ่ม
4. ทำ ตามข้อ 2 และข้อ 3 จนกว่าค่า C ของแต่ละตำแหน่งไม่เปลี่ยน

การใช้งาน K-means ส่วนใหญ่จะใช้สำหรับการทำเหมืองข้อมูล (data mining) และก็ยังมีการใช้ในหลายสาขา เช่น การเรียนรู้ของเครื่อง (ML) การจดจำรูปแบบ (pattern recognition) การวิเคราะห์รูปภาพ (image analysis) การดึงข้อมูล (IR) ชีวสารสนเทศ (bioinformatics) การบีบอัดข้อมูล (data compression) และคอมพิวเตอร์กราฟิก (graphics)

ในปัจจุบันได้มีการนำเทคนิค K-means clustering มาประยุกต์ใช้ในงานวิจัยหลายด้าน อย่างเช่น Ercoli และคณะ [3] ได้นำเสนอวิธีการที่มีประสิทธิภาพสำหรับการดึงข้อมูลคำอธิบายภาพ (visual descriptors retrieval) ด้วยรหัสแฮชโค้ดแบบกะทัดรัด (compact hash codes) โดยใช้ multiple K-means วิธีนี้สามารถใช้กับปัญหาของการค้นหาประมาณค่าเพื่อนบ้านที่ใกล้ที่สุด (Approximate NN search: ANNs) โดยมีการทดสอบกับชุดข้อมูล 3 ชุดข้อมูล คือ 1) BIGANN 2) CIFAR-10 และ 3) MNIST ยังมีงานวิจัยอื่นอีกคือ Arkeman และคณะ [12] ได้มีการปรับใช้ K-meanห ร่วมกับอัลกอริทึมทางพันธุกรรม (multi-objective genetic algorithm) เพื่อเพิ่มประสิทธิภาพ K-means ในข้อมูล Iris และข้อมูล Wine และยังมี Selvakumar และคณะ [13] ได้ประยุกต์ใช้ K-means กับอัลกอริทึมฟัซซี่ซีมีน (Fuzzy C-mean algorithm) ในการตรวจหาระยะ

และรูปร่าง (range and shape) ของเนื้องอกในภาพ MR สมอง อีกงานวิจัยคือ Shan [14] ประยุกต์ใช้งานค่าเอนโทรปีสูงสุด (gray-gradient maximum entropy) เพื่อดึงคุณสมบัติของรูปภาพ และประยุกต์ใช้ K-means เพื่อจำแนกภาพร่วมด้วย และยังมี Kamil และคณะ [15] ใช้ K-means และ Fuzzy C-mean สำหรับการแบ่งส่วนภาพแมมโมแกรม (mammography) ของมะเร็งเต้านม และยังมี Binh และคณะ [16] ได้นำ K-means และ self-organizing map (SOM) โมเดลมาใช้เป็นแนวทางสำหรับการทำการจัดกลุ่มข้อมูล (cluster) โดยประยุกต์ใช้ cluster ใน 2 ระดับ ซึ่งในระดับแรกจะจัดกลุ่มชุดข้อมูลโดยใช้ SOM และในระดับที่สองจะเอาผลลัพธ์ในระดับแรกของ SOM มาจัดกลุ่มโดย K-means และยังประยุกต์ใช้เทคนิค particle swarm optimization-based (PSO) algorithm ร่วมด้วยสำหรับค้นหารัศมีที่เหมาะสม (optimum radius) เพื่อกำหนดจุดศูนย์กลาง (centers) ที่มีประสิทธิภาพ อย่างไรก็ตามเทคนิค K-means แบบดั้งเดิมยังคงเผชิญกับปัญหาความถูกต้องน้อย จึงต้องมีการมาประยุกต์ใช้งานหรือทำงานร่วมกับเทคนิคด้านอื่น ๆ เพื่อเพิ่มประสิทธิภาพ

2.2 ตารางดัชนี (index table)

เทคนิคโครงสร้างดัชนีแบบต้นไม้เดิม (tree-based index structure) เป็นเทคนิคที่นิยมใช้กันอย่างแพร่หลายในการทำดัชนีสำหรับ vector space แต่อย่างไรก็ตามเทคนิคนี้ต้องใช้หน่วยความจำค่อยข้างมาก จึงไม่เหมาะสมและไม่มีประสิทธิสำหรับการใช้งานกับชุดข้อมูลขนาดใหญ่ที่มีมิติสูง (large-scale and high-dimensional datasets) นักวิจัย Jegou และคณะ [5] ได้พัฒนาระบบไฟล์ข้อมูลแบบกลับด้านกับการคำนวณระยะห่างแบบ asymmetric (Inverted file system with asymmetric distance computation: IVFADC) เพื่อรองรับการทำงานกับชุดข้อมูลขนาดใหญ่ระดับพันล้านรายการ (billion-scale datasets) ได้อย่างมีประสิทธิภาพ งานวิจัยดังกล่าวนี้ได้ประยุกต์ใช้เทคนิค K-means สำหรับการเข้ารหัสข้อมูลแบบหยาบ (coarser quantization) และจากนั้นก็ประยุกต์ใช้เทคนิค product quantization (PQ) ด้วยการใช้ residual quantization

Dong และคณะ [17] ประยุกต์ใช้ asymmetric distance computation (ADC) สำหรับการคำนวณระยะห่าง (Euclidean distance) เพื่อค้นหาข้อมูลด้วยตารางดัชนีแบบกลับด้าน (inverted table lookup) โดยที่ความหมายของ asymmetric คือ คำถามหรือคิวรี่กับชุดข้อมูลอ้างอิงอยู่คนละ space (different spaces) กล่าวคือ คำถามเป็นแบบข้อมูลดั้งเดิม (original space) ในขณะที่ชุดข้อมูลอ้างอิงถูกจัดเก็บไว้ในรูปแบบของเลขฐานสอง (binary space) ซึ่งฝั่งของคำถามจะไม่สูญเสีย (quantization loss) ข้อมูลในการเข้ารหัส (quantization) ทำให้ผลการค้นคืนมีความ

ถูกต้องมากกว่าการคำนวณแบบ symmetric distance computation (SDC) วิธีการนี้จะมีการเข้ารหัสทั้งสองฝั่งข้อมูลคือทั้งคำถามและชุดข้อมูลอ้างอิงทำให้มีการสูญเสียข้อมูลขณะทำการเข้ารหัสส่งผลให้ผลลัพธ์มีความแม่นยำน้อยกว่า และยังมีนักวิจัย Gordo และคณะ [18] ได้นำเสนอ ADC บนพื้นฐานของค่าการคาดคะเนทางสถิติ (statistical expectation) และ ค่าขอบล่าง (lower bound) ด้วยค่า Euclidean distance ใน intermediate space เพื่อประมาณค่าของเมตริกซ์ใน original space และยังมี Wang และคณะ [19] ปรับค่าของ ADC ให้เหมาะสม (optimized ADC) ผ่านฟังก์ชันการคำนวณระยะห่าง (distance functions) และส่งผลลัพธ์การเรียนรู้ระยะห่าง (learned distances) ระหว่างคำถามและรหัสไบนารี่

3. การเรียนรู้เชิงลึก (Deep Learning: DL)

Deep Learning (DL) คือวิธีการเรียนรู้แบบอัตโนมัติด้วยซอฟต์แวร์คอมพิวเตอร์ที่เลียนแบบการทำงานของโครงข่ายประสาทสมองมนุษย์ (Neurons) โดยนำระบบโครงข่ายประสาท (Neural Network) มาซ้อนกันหลายชั้น (Layer) โดยที่ชั้นแรกสุดจะทำหน้าที่ในการรับข้อมูลเข้า (Input layer) และชั้นสุดท้ายจะทำหน้าที่ส่งผลลัพธ์การประมวลผลออกมา (Output layer) ส่วนชั้นระหว่างชั้นแรกสุด และชั้นสุดท้าย คือ ชั้นที่ซ่อนอยู่ (Hidden layer) โดย Hidden layer ของแต่ละชั้นจะเปรียบเสมือนเป็นส่วนที่ประกอบด้วย เซลล์ประสาท (neural) จำนวนมาก ซึ่งมีหน้าที่ในการประมวลผลข้อมูลที่รับจากชั้นที่อยู่เหนือกว่า และส่งข้อมูลที่ประมวลผลเสร็จแล้วไปยังชั้นที่อยู่ล่างกว่า ข้อดีของการส่งข้อมูลแบบนี้คือ layer แต่ละ layer สามารถที่จะมีค่าถ่วงน้ำหนัก (weight) ค่าความเอนเอียงของข้อมูล (bias) และ วิธีการประมวลผลทางคณิตศาสตร์ (activation function) ที่เป็นอิสระต่อกันได้ [20]

กระบวนการทำงานของ DL จะทำการเรียนรู้ข้อมูลตัวอย่าง (training set) โดยที่ข้อมูลดังกล่าวจะถูกนำไปใช้ในการตรวจจับรูปแบบ (Pattern) หรือจัดหมวดหมู่ข้อมูล (Classify) โมเดล (model) ของ DL จะให้ความแม่นยำ (accuracy) ที่สูงในการเรียนรู้หลาย ๆ ปัญหา ตั้งแต่การตรวจจับวัตถุ (object detection) ไปจนถึงการรู้จำเสียงพูด (speech recognition) โดยที่ไม่จำเป็นต้องให้ความรู้พื้นฐานใด ๆ ไว้ล่วงหน้า เพียงแค่ให้ข้อมูลตัวอย่าง (input data) โมเดลก็จะทำการเรียนรู้จากข้อมูลและสังเคราะห์เป็นองค์ความรู้ออกมาได้อย่างอัตโนมัติ เช่น การประยุกต์ใช้งานในวงการเกม โดยที่ไม่จำเป็นต้องบอกวิธีเล่น แค่ให้โมเดลเรียนรู้จากผู้เล่นที่เก่ง ๆ เป็นจำนวนมาก ระบบก็เรียนรู้วิธีการเล่นเกมได้อย่างอัตโนมัติ โดยที่การเรียนรู้นั้นเกิดขึ้นจาก 2 เฟส โดยที่เฟสแรกคือ

การประยุกต์ใช้การแปลงแบบไม่เป็นเชิงเส้น (nonlinear transformation) กับข้อมูลที่ได้รับ (input) ได้ผลลัพธ์ (output) ออกมาอยู่ในรูปของโมเดลทางสถิติ (statistical model) เฟสที่สอง คือ การนำโมเดลมาผ่านวิธีการทางคณิตศาสตร์ derivative หรือ การดิฟ โดยทั้ง 2 เฟสนี้จะถูกทำซ้ำจนกว่าจะได้โมเดลที่ความแม่นยำ (accuracy) ในระดับที่น่าพึงพอใจหรือตามค่าที่กำหนดไว้ ซึ่งการทำซ้ำของทั้ง 2 เฟสนี้เรียกว่า iteration และทุกวันนี้ได้มีการนำ DL มาประยุกต์ใช้อย่างแพร่หลาย เช่น รถยนต์ไร้คนขับ (Driverless car) สมาร์ทโฟน search engine ของ google เครื่องจับเท็จ (Fraud detection) โทรทัศน์ และอื่น ๆ อีกมากมาย [20]

งานวิจัยที่เกี่ยวข้องกับปัญหาการวิจัยนี้สามารถแบ่งออกเป็น 2 ส่วนดังนี้

1) ตารางดัชนี (index table)

Chiu และคณะ [21] ได้ประยุกต์ใช้เทคนิค PQ ในการสร้างตารางดัชนี (Product quantization-based indexing) สำหรับการค้นคืนข้อมูลเพื่อนบ้านที่ใกล้ที่สุดโดยการประมาณค่า (ANN search) โดยการใช้ PQ ในการเข้ารหัสข้อมูลระดับสูง (high quantization levels) ด้วยรหัสโค้ดบุ๊ค (compact codebook set) แล้วสร้างตารางดัชนีในแต่ละกลุ่ม (cluster-based index) แล้วพวกเขาได้นำเสนอการเรียนรู้น้ำหนักของความสัมพันธ์ (weighting learning relation) ระหว่างคุณสมบัติคำถาม (query-dependent features) และค่าความสัมพันธ์ของกลุ่มข้อมูล (clusters relevance) จากนั้นค่าความสัมพันธ์ (relevance scores) จะถูกเรียงลำดับความสำคัญโดยค่าของน้ำหนักของคุณสมบัติ (weighted features) ขึ้นตรงต่อคำถาม และต่อมา Song และคณะ [22] ได้เสนอหลายตารางดัชนีกลับด้านแบบใหม่ (inverted multi-index: IMI) โดยโครงสร้างของตารางประยุกต์ใช้จากตารางดัชนีกลับด้านของรหัสไบนารี่ พวกเขาได้ปรับใช้ประสิทธิภาพของโครงสร้าง IMI ด้วยการเลือกตัวแทนบิต 1 ชุดสำหรับประมวลผลบนชุดข้อมูลขนาด 1,000 ล้านรายการ และยังมีอีกหนึ่งงานวิจัย Chiu และคณะ [23] ได้ประยุกต์ใช้ตารางดัชนีแบบกลับด้านเพื่อใช้ในการค้นคืนข้อมูลข้ามประเภท (cross-modal) โดยการเรียนรู้ข้อมูลคำถามจากข้อมูลต้นฉบับกับค่าความน่าจะเป็นหรือความสัมพันธ์ของกลุ่มข้อมูลในการค้นคืนข้อมูลอีกประเภทหนึ่ง อย่างเช่น ใช้ข้อความเป็นคำถามเพื่อค้นหารูปภาพหรือใช้รูปภาพเป็นคำถามเพื่อค้นหาข้อความ เป็นต้น

2) การเรียนรู้เชิงลึก (Deep Neural Network: DNN)

ปัจจุบันได้มีงานวิจัยมากมายที่พัฒนาเพื่อเพิ่มประสิทธิภาพและประยุกต์ใช้ประโยชน์จาก Deep Neural Network (DNN) อย่างเช่น Zhang และคณะ [24] เสนองานวิจัยเรื่อง

Asynchronous stochastic gradient descent for DNN training การใช้งาน DNN ในระบบการเรียนรู้จดจำเสียงพูด (speech recognition systems) เป็นที่รู้จักกันอย่างแพร่หลายว่ามีประสิทธิภาพสูงกว่า conventional GMM-HMM (Gaussian Mixture Model: GMM, Hidden Markov Model: HMM) ทั่วไป แต่จะต้องการเวลาในการฝึกฝน (training) เพราะว่ามีพารามิเตอร์จำนวนมาก การใช้อัลกอริทึมแบบย้อนกลับ (minibatch-based back-propagation: BP) และแบบขนานที่ใช้ในการเรียนรู้ของ DNN นั้นเป็นเรื่องที่ยาก เนื่องจากมีการปรับปรุง (update) โมเดลบ่อยครั้ง งานวิจัยดังกล่าวได้อธิบายถึงวิธีการที่มีประสิทธิภาพเพื่อให้ได้ค่าประมาณของ BP - asynchronous stochastic gradient (ASGD) ซึ่งใช้ในการประมวลผลแบบขนานบน multi-GPU วิธีนี้จะใช้ GPU หลายตัวให้ทำงานแบบอะซิงโครนัสเพื่อคำนวณและปรับปรุงพารามิเตอร์ของโมเดลส่วนกลาง ผลการทดลองแสดงให้เห็นการทำงานเร็วขึ้น 3.2 เท่าบน GPU 4 ตัวเมื่อเทียบกับ GPU ตัวเดียวโดยไม่สูญเสียประสิทธิภาพการจดจำใด ๆ

Snyder และคณะ [25] เสนองานวิจัยเรื่อง X-Vectors: Robust DNN Embeddings for Speaker Recognition โดยใช้วิธีการเสริมข้อมูล (data augmentation) ซึ่ง data augmentation ประกอบด้วยเสียงรบกวน (added noise) และ เสียงก้อง (reverberation) เพื่อเพิ่มประสิทธิภาพของ DNN สำหรับการเรียนรู้จำเสียงพูด (speaker recognition) โดยแบ่งแยกการฝึกฝนระหว่างผู้พูด (speakers) และ จับคู่คำพูดที่ความยาวไม่คงที่ (variable-length utterances) และมีการกำหนดค่ามิติแบบฝังตัวคงที่ (fixed-dimensional) เรียกวิธีการนี้ว่า x-vectors ซึ่งการทดลองพบว่าการฝังตัวใช้ประโยชน์จากชุดข้อมูลการฝึกอบรมขนาดใหญ่ได้ดีกว่าแบบเดิมหรือ baselines (i-vectors)

Zhang และคณะ [26] ได้นำเสนองานวิจัยเรื่อง DNN-based prediction model for spatio-temporal data เนื่องจากความก้าวหน้าทางด้านเทคโนโลยีการสื่อสารแบบไร้สายทำให้ข้อมูล spatio-temporal (ST) มีความพร้อมใช้งานมากขึ้น พวกเขาได้เสนอแบบจำลองการทำนายเรียกว่า DeepST โดยใช้ความรู้เกี่ยวกับโดเมน ST เพื่อออกแบบสถาปัตยกรรมของ DeepST ซึ่งประกอบด้วยสององค์ประกอบ: spatio-temporal และ global โดยที่ spatio-temporal ใช้เฟรมเวิร์คของโครงข่ายประสาทเทียม (CNN) เพื่อสร้างแบบจำลองเชิงพื้นที่ระยะใกล้และระยะไกลพร้อมกัน ส่วนของ global ใช้เพื่อจับปัจจัยทั่วโลกเช่นวันในสัปดาห์วันธรรมดาหรือวันหยุดสุดสัปดาห์ DeepST สามารถสร้างระบบพยากรณ์การไหลของฝูงชนแบบเรียลไทม์ที่เรียกว่า UrbanFlow1 ผลการทดลองแสดงให้เห็นว่า DeepST ทำงานได้ดีกว่าวิธีการพื้นฐาน (baselines)

Juuti และคณะ [27] ได้เสนองานวิจัยเรื่อง PRADA: Protecting Against DNN Model Stealing Attacks พวกเขากล่าวว่าปัจจุบันการนำ ML มาใช้งานแพร่หลายมากขึ้น การปกป้องความลับของโมเดล ML จึงกลายเป็นสิ่งสำคัญยิ่งด้วยเหตุผลสองประการ คือ 1) โมเดลอาจเป็นข้อได้เปรียบทางธุรกิจสำหรับเจ้าของ และ 2) ฝ่ายตรงข้ามอาจใช้ stolen model เพื่อค้นหาตัวอย่างของฝ่ายตรงข้ามที่สามารถถ่ายโอน (transfer or copy) ได้ พวกเขาได้เสนอโมเดลใหม่สำหรับการโจมตีการสกัดที่มีประสิทธิภาพในแง่ของความสามารถในการถ่ายโอนของฝ่ายตรงข้ามที่กำหนดเป้าหมายและไม่กำหนดเป้าหมาย สูงถึง 29 ถึง 44 เปอร์เซ็นต์ และความแม่นยำในการทำนายสูงสุด 46 เปอร์เซ็นต์

Chen และคณะ [28] ได้เสนองานวิจัยเรื่อง Cloud-DNN: An Open Framework for Mapping DNN Models to Cloud FPGAs พวกเขากล่าวว่าเนื่องจาก CNN มีประสิทธิภาพสำหรับแอปพลิเคชันการเรียนรู้ของเครื่อง (ML) ที่หลากหลาย แต่ก็มีความซับซ้อน (complex) ในการคำนวณสูง ทำให้เกิดความท้าทายที่สำคัญต่อการนำไปใช้ในวงกว้างในการทำงานแบบเรียลไทม์และการประหยัดพลังงาน เทคนิค FPGA มีบทบาทสำคัญสำหรับการคำนวณของ CNN ที่มีประสิทธิภาพสูงและประหยัดพลังงานสำหรับทั้งอุปกรณ์เคลื่อนที่ (เช่น UAV รถยนต์ที่ขับเคลื่อนด้วยตนเองและอุปกรณ์ IoT) และคลาวด์คอมพิวติ้ง อย่างไรก็ตามในการนำระบบ CNN ไปใช้กับ FPGA อย่างมีประสิทธิภาพยังคงเป็นปัญหา FPGA บนคลาวด์ในปัจจุบันยังมีข้อจำกัด ด้านการออกแบบในทางสถาปัตยกรรมยิ่งเพิ่มความท้าทาย เพื่อจัดการกับความท้าทายเหล่านี้ พวกเขาเสนอเครื่องมืออัตโนมัติแบบโอเพนซอร์สที่เรียกว่า Cloud-DNN โดยใช้โมเดล CNN ที่ได้รับการฝึกฝนใน Caffe เป็นข้อมูลนำเข้า (input) เพื่อดำเนินการของการเปลี่ยนแปลงและแมปโมเดลกับ FPGA บนคลาวด์ Cloud-DNN สามารถปรับปรุงประสิทธิภาพการออกแบบโดยรวมของ CNN บน FPGA ได้อย่างมาก ในขณะที่ตอบสนองความต้องการด้านการคำนวณที่เกิดขึ้นใหม่ การออกแบบนี้เป็นอีกทางเลือกเมื่อเทียบกับตัวเลือกบนคลาวด์อื่น ๆ (เช่น GPU หรือ TPU) ผลการทดลองแสดงให้เห็นถึงการปรับปรุงประสิทธิภาพสูงสุด 104.55 เท่าเมื่อเทียบกับการใช้งาน CPU และความสามารถในการใช้งานมีความยืดหยุ่นและคุณภาพที่ดีเทียบได้กับการใช้การ DNN อื่น ๆ บน FPGA แบบเดี่ยว (stand-alone)

Yang และคณะ [29] เสนองานวิจัยเรื่อง A Survey of DNN Methods for Blind Image Quality Assessment ซึ่งก็คือวิธีการประเมินคุณภาพของภาพคนตาบอด (Blind image quality assessment: BIQA) โดยมีจุดมุ่งหมายเพื่อทำนายคุณภาพของภาพตามที่มนุษย์รับรู้โดยไม่ต้องเข้าถึง

ภาพอ้างอิง วิธีการเรียนรู้เชิงลึก (DL) ได้รับความสนใจอย่างมากในการวิจัยและมีการพิสูจน์แล้วว่ามีประโยชน์สำหรับ BIQA ทีมงานนักวิจัยนี้ได้จัดเตรียมแบบสำรวจที่ครอบคลุมวิธีการ DNN ต่างๆ สำหรับ BIQA ขั้นแรกวิเคราะห์วิธีการประเมินคุณภาพตาม DNN ที่มีอยู่อย่างเป็นระบบตามบทบาทของ DNN จากนั้นจะเปรียบเทียบประสิทธิภาพการทำนายของวิธี DNN ต่าง ๆ บนฐานข้อมูลสังเคราะห์ (LIVE, TID2013, CSIQ, LIVE) และฐานข้อมูลที่แท้จริง (LIVE challenge) โดยให้ข้อมูลสำคัญที่สามารถช่วยให้เข้าใจคุณสมบัติพื้นฐานระหว่างวิธี DNN ที่แตกต่างกันสำหรับ BIQA สุดท้ายพวกเขาได้อธิบายถึงความท้าทายที่เกิดขึ้นใหม่ในการออกแบบและฝึกอบรม BIQA ที่ใช้ DNN พร้อมกับแนวทางบางประการที่ควรค่าแก่การตรวจสอบเพิ่มเติมในอนาคต

Narayanan และคณะ [30] ได้เสนองานวิจัยเรื่อง PipeDream: generalized pipeline parallelism for DNN training พวกเขากล่าวว่าเนื่องจากการฝึกอบรม DNN ใช้เวลานานมาก (time-consuming) และจำเป็นต้องมีการขนานตัวเร่งความเร็วหลายตัว (multi-accelerator) ที่มีประสิทธิภาพ แนวทางปัจจุบันในการฝึกอบรมแบบขนานส่วนใหญ่จะใช้การขนานกันภายในชุดการทำงานซึ่งการฝึกอบรมซ้ำเพียงครั้งเดียวจะแยกออกจากงานที่มีกำลังรันอยู่ แต่ได้รับผลตอบแทนลดลงเมื่อมีงานจำนวนที่สูงขึ้น พวกเขาได้เสนอเทคนิค PipeDream ซึ่งเป็นระบบที่เพิ่มการไปป์ไลน์ระหว่างแบตช์ให้กับการขนานกันภายในแบทช์เพื่อปรับปรุงทรูพุตการฝึกอบรมแบบคู่ขนานให้ดียิ่งขึ้น ซึ่งจะช่วยให้สามารถคำนวณทับซ้อนกับการสื่อสารได้ดีขึ้นและลดปริมาณการสื่อสาร ซึ่งแตกต่างจากการไปป์ไลน์แบบเดิม การฝึก DNN เป็นแบบสองทิศทางโดยที่การส่งต่อไปข้างหน้าและการส่งย้อนกลับ การไปป์ไลน์อย่างเดียวอาจส่งผลให้เกิดความไม่ตรงกันในเวอร์ชันที่ใช้ในการส่งต่อและย้อนกลับหรือไปป์ไลน์มากเกินไปและประสิทธิภาพของฮาร์ดแวร์ลดลง เพื่อจัดการกับความท้าทายเหล่านี้ PipeDream จะสร้างโมเดลพารามิเตอร์สำหรับการคำนวณและกำหนดเวลาเดินหน้าและถอยหลังของมินิแบทช์ที่แตกต่างกันไปพร้อมกันกับการทำงานที่แตกต่างกันโดยมีแผงวางท่อน้อยที่สุด PipeDream ยังแบ่งชั้น DNN ระหว่างการงานโดยอัตโนมัติเพื่อสร้างสมดุลระหว่างการทำงานและลดการสื่อสาร การทดลองกับการทำงานของ DNN แบบจำลองและการกำหนดค่าฮาร์ดแวร์แสดงให้เห็นว่า PipeDream ฝึกโมเดลให้มีความแม่นยำสูงถึง 5.3 เท่าเร็วกว่าเทคนิคการขนานภายในแบทช์ที่ใช้กันทั่วไป

3) การค้นคืนข้ามมีเดีย (Cross-media retrieval)

Peng และคณะ [9] เสนอว่าการค้นคืนข้อมูลข้ามมีเดีย (cross-media retrieval) เป็นเทคนิคที่น่าสนใจและมีความท้าทายเป็นอย่างยิ่งในปัจจุบัน เพราะสื่อทางด้านมัลติมีเดียได้เพิ่มขึ้นอย่างก้าวกระโดดวันต่อวัน การสืบค้นข้อมูลแบบข้ามมีเดียจึงเป็นการเพิ่มขีดความสามารถและความยืดหยุ่นในการค้นคืนข้อมูลด้านมัลติมีเดีย อย่างเช่น สามารถสืบค้นข้อมูลรูปภาพโดยใช้ข้อความเป็นคำค้น หรือสามารถสืบค้นข้อความโดยใช้รูปภาพเป็นคำค้น หรือสามารถใช้คำค้นประเภทหนึ่งแล้วผลการสืบค้นแสดงในลักษณะสื่ออื่น ๆ เป็นต้น งานวิจัยดังกล่าวได้สร้างชุดข้อมูลใหม่ชื่อ XMedia ซึ่งเป็นชุดข้อมูลที่เปิดเผยต่อสาธารณะชุดแรกโดยมีสื่อถึง 5 ประเภท คือ ข้อความรูปภาพ วิดีโอ เสียง และแบบจำลอง 3 มิติ เพื่อจะดึงดูดให้นักวิจัยให้มุ่งเน้นไปที่การดึงข้อมูลข้ามมีเดีย และยังมีงานวิจัยของ Li และคณะ [31] เสนอการแฮชแบบไม่ต่อเนื่อง (discrete) สำหรับการค้นคืนข้ามมีเดีย (cross-view) แบบไม่มีลาเบล (unsupervised) แบบออนไลน์สำหรับขนาดใหญ่ งานวิจัยของ Sharma และคณะ [32] นำเสนอการทำดัชนีในการค้นคืนรูปภาพ โดยการเข้าแฮชแบบ unsupervised ข้อมูลขนาดใหญ่แบบหลายคุณสมบัติ (multi-feature) Zhang และคณะ [33] ได้นำเสนอเทคนิค Multiple hash codes joint learning method (MOON) สำหรับการดึงข้อมูลข้ามสื่อ โดยจะเรียนรู้รหัสแฮชที่มีความยาวหลายส่วน (Multiple hash codes) พร้อมกัน เพื่อปรับปรุงวิธีการพื้นฐานโดยไม่ต้องฝึกอบรมใหม่ในการดึงข้อมูล

บทที่ 3
วิธีดำเนินการวิจัย

3.1 รูปแบบการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง (Experimental research) โดยใช้ชุดข้อมูลมัลติมีเดีย 4 ชุดข้อมูล วิธีการทดสอบการค้นคืนแบบดั้งเดิมหรือการค้นหาแบบละเอียด และโมเดลทำนายผ่านตารางดัชนีแบบกลับด้าน เพื่อเปรียบเทียบประสิทธิภาพของแต่ละระบบ รูปแบบการเปรียบเทียบศึกษาการวิจัยแสดงดังตารางที่ 3.1

ตารางที่ 3.1 รูปแบบการเปรียบเทียบศึกษาการวิจัย

ผลการค้นคืนแบบละเอียด	โมเดลการทำนาย (Prediction model)	ผลการค้นคืน ผ่านโมเดลการทำนาย	
O_1	X	O_2	

กำหนดให้ O_1 คือ ผลการค้นคืนข้อมูลแบบดั้งเดิมหรือแบบละเอียด

X คือ โมเดลการทำนาย

O_2 คือ ผลการค้นคืนด้วยโมเดลการทำนายผ่านตารางดัชนีแบบกลับด้าน

โดยการดำเนินการวิจัยจะทำการทดลองและเปรียบเทียบผลการค้นคืนข้อมูลของค่า O_1 และ O_2 เพื่อวัดประสิทธิภาพของแต่ละโมเดล X

3.2 ประชากรและกลุ่มตัวอย่าง

งานวิจัยนี้ใช้ชุดข้อมูลมัลติมีเดียมาตรฐานที่เป็นที่นิยม 4 ชุดข้อมูลในการดำเนินการทดลองและประเมินผลการวิจัย ได้แก่ 1) ชุดข้อมูล Wiki [6] 2) ชุดข้อมูล MIRFlickr [7] 3) ชุดข้อมูล NUS-WIDE [8] และ 4) ชุดข้อมูล XMedia [9] โดยแต่จะชุดข้อมูลประกอบด้วย ข้อมูลภาพ ข้อความ และที่น่าท้าทายคือในส่วนของชุดข้อมูล XMedia จะมีทั้งหมด 5 ประเภทโดยมีข้อมูลเพิ่มอีก 3 ประเภท คือ ข้อมูลเสียง (audio) ข้อมูลวิดีโอ (video) และข้อมูล 3 มิติ (3D)

ตารางที่ 3.2 รายละเอียดชุดข้อมูล ชุดข้อมูล WIKI ชุดข้อมูล MIRFlickr และ ชุดข้อมูล NUS-WIDE

ชุดข้อมูล	WIKI	MIRFlickr	NUS-WIDE
ข้อมูลทั้งหมด (reference set)	2866	15902	184711
ข้อมูลสำหรับเรียนรู้ (training set)	2173	10000	20000
ข้อมูลที่ใช้ทดสอบ (testing set)	693	836	1866
จำนวนประเภทข้อมูล (label)	10	24	10

ตารางที่ 3.3 รายละเอียดชุดข้อมูล XMedia

ประเภทข้อมูล (media)	รูปภาพ	ข้อความ	ข้อมูล 3 มิติ	วิดีโอ	เสียง
ข้อมูลทั้งหมด	5000	5000	500	1143	1000
ข้อมูลสำหรับการเรียนรู้	4000	4000	400	969	800
ข้อมูลที่ใช้ทดสอบ	1000	1000	100	174	200
จำนวนประเภทข้อมูล	20	20	20	20	20

ชุดข้อมูลทุกชุดข้อมูลจะถูกสุ่มตัวอย่างเพื่อแบ่งออกเป็นชุดข้อมูลที่ใช้ในกาเรียนรู้ของโมเดล (training set) และข้อมูลที่ใช้ในการทดสอบประสิทธิภาพของโมเดล (testing set) โดยมีรายละเอียดการแบ่งข้อมูลแสดงไว้ดังในตารางที่ 3.2 (รายละเอียดชุดข้อมูล Wiki ชุดข้อมูล MIRFlickr และ ชุดข้อมูล NUS-WIDE) และ ตารางที่ 3.3 (รายละเอียดชุดข้อมูล XMedia) ชุดข้อมูลจะถูกนำไปจัดกลุ่มด้วยอัลกอริทึม K-means clustering เพื่อสร้างเป็นรหัสไบนารี่ ชุดข้อมูล Wiki [6] คัดเลือกมาจาก Wiki pedia’s editors ตั้งแต่ปี 2552 จำนวน 2700 บทความ (Feature articles) และได้จัดกลุ่มข้อมูลไว้ 29 กลุ่ม (Classes) ข้อมูลในชุดข้อมูล Wiki pedia จะเป็นข้อมูลที่เป็นที่นิยมหรือน่าสนใจมากที่สุด 10 รายการ สุดท้ายแล้วชุดข้อมูลนี้ประกอบด้วย 2866 รายการ เราทำการสุ่มเพื่อแยกข้อมูลออกเป็นข้อมูลชุดข้อมูลสำหรับการเรียนรู้ทั้งหมด 2173 รายการ และข้อมูลสำหรับใช้ทดสอบประสิทธิภาพโมเดล 693 รายการ โดยที่ข้อมูลทั้งหมดประกอบด้วยคู่ภาพและข้อความ (Text-image pairs) พร้อมด้วยระบุประเภทของข้อมูลหรือลาเบล (Label) ของ 10 กลุ่มข้อมูลนั้น รายละเอียดของ

รูปภาพแต่ละรูปถูกนำเสนอในรูปแบบของ Bag-of-visual word (BoVW) histogram ของ 128-codeword SIFT codebook รายละเอียดของข้อความแต่ละข้อความถูกอธิบายด้วย Histogram of a 10-topic LDA model สำหรับข้อมูลเฉลยเพื่อนบ้านที่ใกล้เคียง (Ground truth neighbors) คือ ชุดคู่ข้อมูลที่มี label เดียวกัน

ชุดข้อมูล MIRFlickr [7] มีข้อมูลเดิมทั้งหมด 25000 ตัวอย่าง (Instances) ถูกรวบรวมจากเว็บไซต์ Flickr โดยแต่ละตัวอย่างประกอบด้วยรูปภาพที่สัมพันธ์กับข้อความที่เกี่ยวข้อง (Textual tag) โดยที่มีการกำหนดลาเบลไว้ล่วงหน้า (Predefined semantic labels) ตั้งแต่ 1 ถึง 24 (Predefined semantic labels) จากนั้นเราลบเฉพาะรายการที่ไม่มี Textual tags หรือมี Textual tags น้อยกว่า 20 ครั้งออก โดยแต่ละรูปภาพจำนำเสนอด้วยตัวอักษรของ 150-D edge histogram และแต่ละข้อความถูกเสนอด้วยการใช้ PCA บน Binary tagging และแสดงในรูปแบบ Feature vector of 500-D เราทำการสุ่มข้อมูลจำนวน 2000 รูปภาพจากข้อมูลทั้งหมดเพื่อใช้เป็นข้อมูลในการทดสอบหรือข้อมูลคำถาม (Query set) และที่เหลือจะเป็นฐานข้อมูลหรือชุดข้อมูลที่ใช้ในสำหรับการสืบค้น (reference set) และสุ่มข้อมูล 10000 รายการจาก Reference set เพื่อใช้ในการเรียนรู้ของโมเดล (Training set) ข้อมูลเพื่อนบ้านใกล้เคียงหรือคำตอบ คือคู่รูปภาพหรือข้อความที่มีการใช้ Label ร่วมกันอย่างน้อย 1 รายการ

ชุดข้อมูล NUS-WIDE [8] เดิมมีข้อมูล 260648 รายการ แต่ละรายการประกอบด้วยรูปภาพที่มีการกำหนดลาเบลไว้อย่างน้อย 1 ลาเบลจากทั้งหมด 81 ลาเบล เราเลือกลาเบลที่มีความถี่สูงสุด 10 รายการแรก ดังนั้นข้อมูลที่เหลือทั้งหมดคือ 184711 รายการของคู่รูปภาพและข้อความ โดยข้อมูลรูปภาพแต่ละรูปถูกนำเสนอด้วย 500 มิติ (Dimensional) BoVW vector จาก hand-crafted feature และข้อความแต่ละข้อความถูกนำเสนอด้วย 1000-dimensional bag-of-words (BoW) vector จากนั้นเราทำการสุ่มข้อมูลเพื่อใช้เป็นข้อมูลที่ใช้ในการทดสอบโมเดลจำนวนร้อยละ 5 ของข้อมูลทั้งหมด คือ 1866 รายการ และในส่วนของข้อมูลที่เหลือจะเป็น Reference set สำหรับข้อมูลที่ใช้ในการสอนโมเดลเราทำการสุ่มจำนวน 20000 รายการจาก Reference set ข้อมูลเพื่อนบ้านที่ใกล้เคียงหรือคำตอบ คือคู่รูปภาพหรือข้อความที่มีการใช้ Label ร่วมกันอย่างน้อย 1 รายการ

ชุดข้อมูล XMedia [9] ข้อมูลทั้งหมดถูกรวบรวมจากเว็บไซต์บนอินเตอร์เน็ตจากหลายแหล่งดังนี้ Wiki pedia YouTube Flickr Freedsound Findsound 3D Warehouse และ Princeton 3D Model Search Engine ข้อมูลทั้งหมดประกอบด้วย 5 ประเภทดังนี้ 1) ข้อความ 5000 รายการ

2) รูปภาพ 5000 รายการ 3) วิดีโอ 1143 รายการ 4) เสียง 1000 รายการ และ 5) ข้อมูล 3 มิติ 500 รายการ ข้อมูลทั้งหมดแบ่งออกเป็น 20 กลุ่มข้อมูล เราทำการสุ่มข้อมูลแต่ละประเภทเพื่อแยกสำหรับการใช้ในการสอนเพื่อการเรียนรู้ของโมเดลและข้อมูลเพื่อใช้ในการทดสอบประสิทธิภาพของโมเดลดังนี้ สำหรับข้อมูลรูปภาพและข้อความจะแบ่งเป็นข้อมูลเพื่อใช้สอนโมเดลจำนวน 4000 รายการ และใช้เพื่อทดสอบโมเดล 1000 รายการ ข้อมูลวิดีโอจะแบ่งเพื่อใช้ในการทดสอบโมเดล 174 รายการ ในส่วนที่เหลือใช้เพื่อการสอนโมเดล ข้อมูล 3 มิติสุ่มมา 100 รายการเพื่อใช้ในการทดสอบระบบ ในส่วนที่เหลือใช้เพื่อสอนโมเดลจำนวน 400 รายการ ข้อมูลเสียงสุ่มมาจำนวน 200 รายการเพื่อใช้ในการทดสอบ และในส่วนที่เหลือใช้เพื่อการสอนโมเดลจำนวน 800 รายการ สำหรับรายละเอียดของการเสนอข้อมูลรูปภาพเราใช้ BoVW histogram ของ 128-codeword SIFT codebook ในส่วนของข้อความแต่ละข้อความแสดงโดย Histogram of a 10-topic LDA model ในส่วนของคลิปเสียงแต่ละคลิปแสดงโดย 29-dimensional MFCC feature สำหรับวิดีโอเริ่มต้นด้วยกันแบ่งวิดีโอออกเป็น Video shots จากนั้นก็ทำ Extracted เป็น 128-dimensional BoVW histogram feature สำหรับวิดีโอแต่ละ Keyframe สุดท้ายข้อมูล 3 มิติจะนำเสนอด้วย concatenated 4700-dimensional vector of the LightField descriptor set ในส่วนของข้อมูลเพื่อนบ้านใกล้เคียงหรือคำตอบ คือคู่ข้อมูลที่ใช้ลาเบลเดียวกัน

3.3 เครื่องมือที่ใช้ในการรวบรวมข้อมูล

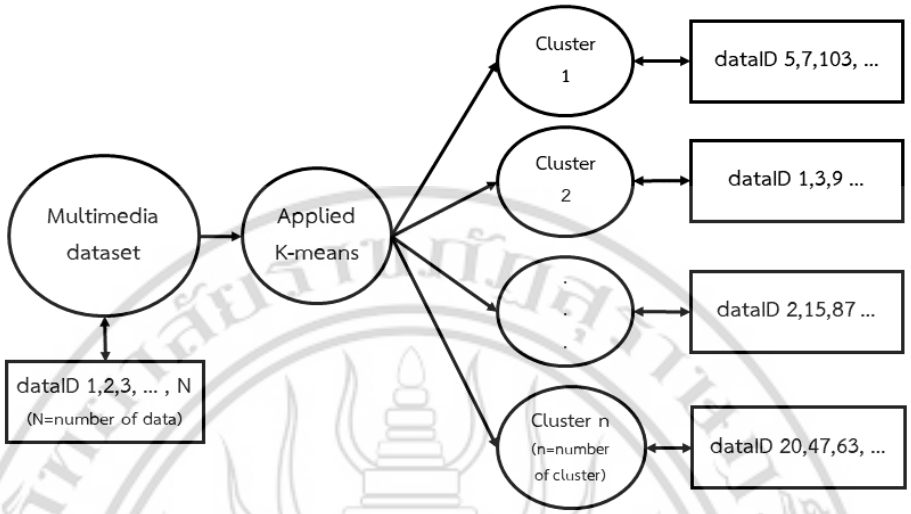
ผู้วิจัยศึกษาข้อมูลจากเอกสารหรืองานวิจัยที่เกี่ยวข้องที่ใช้ชุดข้อมูลมัลติมีเดีย โดยอ้างอิงจากงานวิจัยเรื่อง An overview of cross-media retrieval: concepts, methodologies, benchmarks, and challenges [1] ซึ่งสามารถสรุปได้ดังตารางที่ 3.4 โดยมีรายละเอียดดังนี้ 1) ชุดข้อมูล Wiki มีการอ้างถึงหรือใช้เพื่อทดลองในงานวิจัยจำนวน 38 รายการ 2) ชุดข้อมูล NUS-WIDE จำนวน 30 รายการ 3) ชุดข้อมูล Pascal VOC 2007 จำนวน 9 รายการ 4) ชุดข้อมูล Pascal Sentence จำนวน 9 รายการ 5) ชุดข้อมูล MIRFlickr จำนวน 7 รายการ 6) ชุดข้อมูล Corel จำนวน 6 รายการ 7) ชุดข้อมูล LabelMe จำนวน 5 รายการ และ 8) ชุดข้อมูล XMedia จำนวน 3 รายการ และในการทดสอบประสิทธิภาพของงานวิจัยนี้เราได้เลือกและรวบรวมชุดข้อมูลมัลติมีเดียที่เป็นที่นิยมและใช้กันอย่างแพร่หลาย (Benchmark) จำนวน 4 ชุดข้อมูล ได้แก่ 1) ชุดข้อมูล Wiki [6] 2) ชุดข้อมูล MIRFlickr [7] 3) ชุดข้อมูล NUS-WIDE [8] และ 4) ชุดข้อมูล XMedia [9]

ตารางที่ 3.4 สรุปรายการชุดข้อมูลมัลติมีเดียที่นิยมใช้กันอย่างแพร่หลาย

ลำดับ	ประเภทข้อมูล (media)	รูปภาพ
1	Wiki	38
2	NUS-WIDE	30
3	Pascal VOC 2007	9
4	Pascal Sentence	9
5	MIRFlickr	7
6	Corel	6
7	LabelMe	5
8	XMedia	3

3.4 วิธีการเก็บรวบรวมข้อมูล

หลังจากศึกษาข้อมูลจากเอกสารและงานวิจัยที่เกี่ยวข้องแล้ว นักวิจัยได้ทำการดาวน์โหลดชุดข้อมูลมัลติมีเดียจากเว็บไซต์ที่ให้บริการชุดข้อมูลดังกล่าวจากเว็บไซต์ที่ให้บริการ แล้วจึงจัดเตรียมข้อมูล (Data preparing) ตามรายละเอียดดังหัวข้อ 3.2 เพื่อใช้ในการดำเนินงานวิจัยตามลำดับ โดยนำข้อมูลมัลติมีเดียของแต่ละชุดมาจัดกลุ่มด้วยเทคนิค K-means เพื่อสร้างตารางดัชนีแบบกลับด้าน ซึ่งสามารถอธิบายขั้นตอนดังภาพตัวอย่างที่ 3.1 ขั้นตอนแรกเป็นการจัดเตรียมชุดข้อมูลโดยมีข้อมูลจำนวน *N* รายการ ขึ้นอยู่กับชุดข้อมูล ขั้นตอนที่สองเป็นการจัดกลุ่มด้วยเทคนิค K-means จำนวนกลุ่ม (Cluster) ขึ้นอยู่กับค่า *n* ซึ่งเป็นค่าที่สามารถกำหนดได้ว่าจะแบ่งข้อมูลออกเป็นกี่กลุ่ม ขั้นตอนสุดท้ายคือการนำข้อมูลในแต่ละกลุ่มมาสร้างตารางดัชนี โดยใช้หมายเลขกลุ่มเป็นรหัสที่ใช้แทนชุดข้อมูลที่อยู่ในกลุ่มนั้น ๆ



ภาพที่ 3.1 การจัดกลุ่มและสร้างดัชนีข้อมูล

3.5 สถิติที่ใช้ในการวิเคราะห์ข้อมูล

วิธีการประเมินประสิทธิภาพการค้นคืนข้อมูลแบบพื้นฐานและเป็นที่นิยมใช้กันอย่างแพร่หลายประกอบด้วย ค่าพรีซิชั่น (Precision) รีคอล (Recall) และ ค่าเฉลี่ยพรีซิชั่น (MAP) โดยที่ค่าพรีซิชั่นจะบอกถึงประสิทธิภาพของระบบในการค้นคืนเอกสารโดยดูจากอัตราส่วนของจำนวนเอกสารที่ถูกต้องจากเอกสารที่ถูกเลือกมาทั้งหมด ส่วนรีคอลหมายถึงประสิทธิภาพของการค้นคืนเอกสารโดยดูจากอัตราส่วนจำนวนเอกสารที่ถูกต้องที่เลือกมาต่อจำนวนเอกสารที่ถูกต้องทั้งหมดที่อยู่ในคอลเล็คชั่น (Collection) หากต้องการจะวัดค่าประสิทธิภาพของ Model จำเป็นต้องได้ค่ามา 1 ค่าเพื่อเป็นตัวแทนประสิทธิภาพของ Model โดยปกติแล้วจะใช้ค่าของ Average precision เป็นตัวแทนของประสิทธิภาพของ Model โดยที่ Average Precision คำนวณได้จริงจากพื้นที่ใต้กราฟของ Precision กับ Recall โดยจะมีค่าอยู่ระหว่าง 0 ถึง 1 ถ้าเป็น 1 แสดงว่าโมเดลทำงานได้ดีเยี่ยม กราฟของ Precision จะเป็นสี่เหลี่ยม คือมี Precision เป็น 1 และ Recall เป็น 1 และยังมีค่ามาตรฐานอีกอันหนึ่งที่ใช้ในการประเมินประสิทธิภาพระบบค้นคืนสารสนเทศคือ Mean Average Precision (MAP) ซึ่งจะให้ค่าเฉลี่ยของพรีซิชั่นจากทั้งหมดของทุกๆ ข้อคำถาม วิธีการนี้จะทำการหาค่าพรีซิชั่น ณ ทุก ๆ ตำแหน่งที่ปรากฏเอกสารที่เกี่ยวข้อง (Relevance) ใน Top *k* ของรายการผลลัพธ์หลังจากนั้นนำค่าพรีซิชั่นเหล่านั้นมาหาค่าเฉลี่ยซึ่งจะเรียกว่า ค่าเฉลี่ยพรีซิชั่น โดยมีสูตรการคำนวณดังนี้ [34]

Precision = TP/(TP + FP) (3.1)

Recall = TP/(TP + FN) (3.2)

โดยที่

- True Positive (TP) คือ สิ่งที่ทาย (prediction) ว่ามันมีจริง และ เฉลย (Ground Truth) บอกว่ามีจริง

- True Negative (TN) คือ สิ่งที่ทายว่ามันไม่มีจริง และเฉลยก็บอกว่ามันไม่มีจริง

- False Positive (FP) คือ สิ่งที่ทายว่ามันมีจริง แต่เฉลยก็บอกว่ามันไม่จริง (ทำนายผิด)

- False Negative (FN) คือ สิ่งที่ทายว่ามันมีไม่จริง แต่เฉลยก็บอกว่ามันมีจริง (ทำนายผิด)

การคำนวณ MAP [14] ใช้ในการประเมินความถูกต้องของผลการค้นคืนข้อมูลจากชุดข้อมูลคำถาม (queries) Q ดังสมการที่ 3.2:

$$MAP = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{R}\sum_{j=1}^{R} pr(j) \cdot rel(j), \quad (3.3)$$

โดยที่ Q คือจำนวนคำถาม

R คือจำนวนเอกสารที่ค้นคืน

pr(j) หมายถึงความแม่นยำของการค้นคืนข้อมูล

rel(j) = 1 หากเอกสารที่ค้นคืนของลำดับที่ *jth* มีความเกี่ยวข้องกับคำถาม ถ้าไม่เกี่ยวข้อง *rel(j)* = 0 เอกสารที่เกี่ยวข้องถูกกำหนดให้เป็นคู่ที่มีป้ายลาเบล (label) ร่วมกันอย่างน้อยหนึ่งป้ายหรือมีป้ายลาเบลเดียวกัน การคำนวณ MAP คือค่าเฉลี่ยของความแม่นยำของคำค้นหรือคำถามทั้งหมด โดยที่ค่า MAP ที่สูงกว่าคือมีประสิทธิภาพการค้นคืนข้อมูลที่ดีกว่า

3.6 ขั้นตอนการทดลอง

ภาพที่ 3.2 แสดงกรอบ (framework) ของการค้นคืนข้อมูลของกระบวนการวิธีที่นำเสนอ

จากภาพที่ 3.2 แสดงกรอบของการค้นคืนข้อมูลเพื่อนบ้านที่ใกล้ที่สุดของคำถาม โดยกระบวนการทำงานประกอบด้วย 2 ส่วน คือ 1) ส่วนของการสอนโมเดล (Training part) ขั้นตอนที่ 1-6 และ 2) ส่วนของการค้นคืน (Searching part) และวัดประสิทธิภาพในขั้นตอนที่ 7-11 โดยขั้นตอนการทำงานในส่วนที่หนึ่ง มีลำดับขั้นตอนการทำงานดังนี้

1.1) ระบบรับข้อมูลมัลติมีเดียซึ่งเป็นข้อมูลดิบประกอบด้วย ข้อความ รูป เสียง วิดีโอ และ 3 มิติ (Input raw data หรือ Data feature)

1.2) จัดกลุ่มข้อมูลด้วยการประยุกต์ใช้อัลกอริทึม K-means clustering และนำรหัสกลุ่มข้อมูลมาสร้างเป็นตัวแทนข้อมูลในกลุ่มของแต่ละประเภทข้อมูล (Media) ในแต่ละชุดข้อมูล (Datasets)

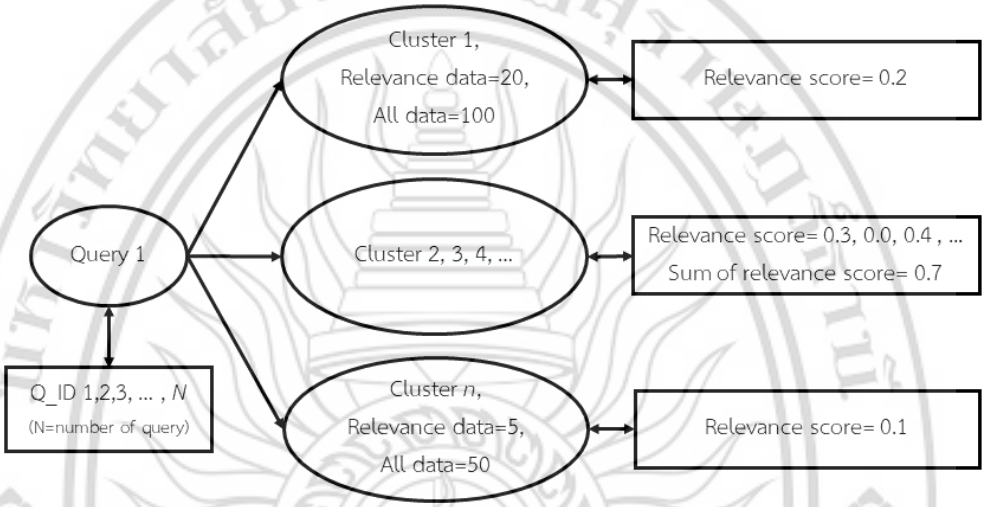
1.3) แบ่งข้อมูลที่ใช้สอนโมเดล (Training data) และทดสอบ (Test data)

1.4) สร้างตารางดัชนีแบบกลับด้านจากรหัสกลุ่มข้อมูล K-means

1.5) คำนวณค่าความสัมพันธ์หรือความน่าจะเป็นระหว่างคำถามกับคำตอบในแต่ละกลุ่มข้อมูล (Cluster) สามารถดูตัวอย่างการคำนวณดังภาพที่ 3.3 โดยใช้สูตรการคำนวณดังต่อไปนี้

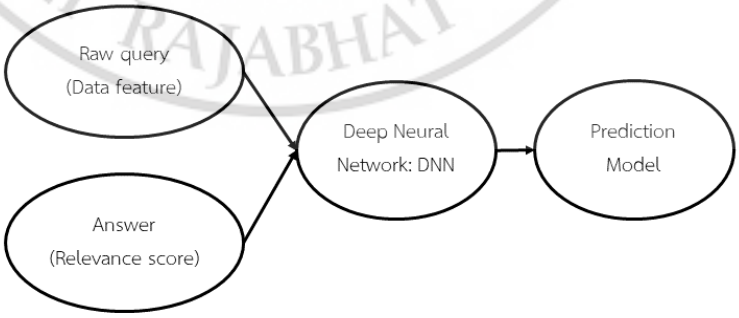
$$\text{ความสัมพันธ์} = \frac{\text{จำนวนข้อมูลที่มีความสัมพันธ์กับคำถามในแต่ละกลุ่มข้อมูล}}{\text{จำนวนข้อมูลทั้งหมดในแต่ละกลุ่มข้อมูล}} \qquad (3.4)$$

สมมติคำถามที่ 1 (Query 1) มีข้อมูลที่มีความสัมพันธ์กับกลุ่มข้อมูลที่ 1 (Cluster 1) จำนวน 20 ข้อมูล และข้อมูลทั้งหมดใน Cluster นั้น มี 100 ข้อมูลจะสามารถำนวณความสัมพันธ์ได้ 20/100 = 0.2 โดยคำนวณความสัมพันธ์ทุกกลุ่มข้อมูล (*n* คือจำนวน Cluster ที่เลือกตอนทำ K-means) โดยค่ารวมของความสัมพันธ์ทั้งหมดจะมีค่าเท่ากับ 1



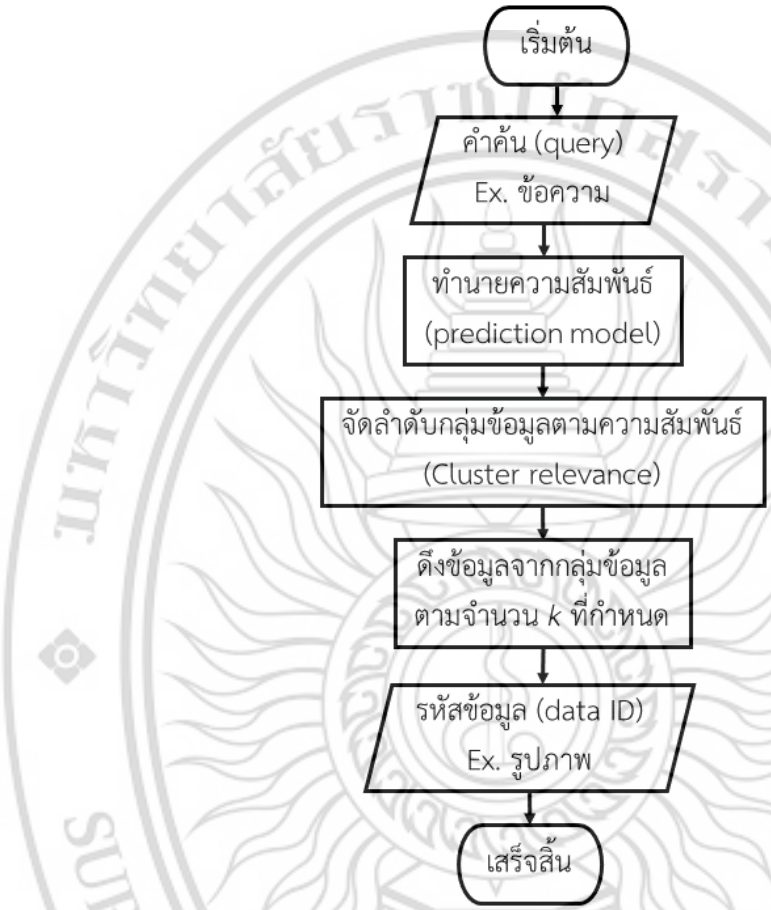
ภาพที่ 3.3 การคำนวณคะแนนความสัมพันธ์

1.6) จากภาพที่ 3.4 แสดงขั้นตอนการสร้างโมเดลการทำนาย (Prediction model) โดยขั้นตอนการเรียนรู้คุณสมบัติข้อมูล (Features) ของข้อมูลคำถามจับคู่กับความสัมพันธ์ของข้อมูล (Relevance score) ของแต่ละกลุ่มข้อมูล (Cluster) เพื่อสร้างโมเดลการเรียนรู้ ที่ใช้ในการทำนายข้อมูลคำถามใหม่ที่ยังไม่ได้ผ่านการเรียนรู้มาก่อนว่าจัดอยู่ในกลุ่มใด

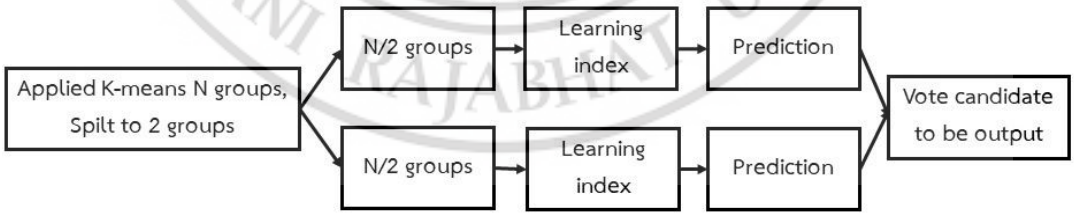


ภาพที่ 3.4 โมเดลการทำนาย

ในส่วนที่ 2 ขั้นตอนของการค้นคืนข้อมูล แสดงตัวอย่างดังภาพที่ 3.5 โดยมีขั้นตอนการทำงานดังนี้



ภาพที่ 3.5 ขั้นตอนการค้นคืนข้อมูล



ภาพที่ 3.6 การประยุกต์ใช้หลายดัชนี K-means (Multi-K-means)

2.1) ป้อนคำถามใหม่เพื่อทดสอบโมเดลการเรียนรู้โดยส่งคำถามของสื่อประเภทหนึ่ง (เช่น ข้อความ) เข้าโมเดลดังกล่าว

2.2) โมเดลเรียนรู้คุณสมบัติข้อมูลแล้วจึงทำนายความสัมพันธ์ข้อมูล (Cluster relevance scores) เพื่อเรียงลำดับความเกี่ยวข้องกันของกลุ่มข้อมูลอีกประเภทหนึ่ง (เช่น รูปภาพ) จากความสัมพันธ์มากที่สุดไปยังน้อยที่สุด

2.3) จัดลำดับความสัมพันธ์โดยเลือกกลุ่มข้อมูลที่มีความสูงสุดก่อน

2.4) ดึงข้อมูลตัวแทนคำตอบ (Candidate) จากกลุ่มข้อมูลอันดับต้น ๆ (Top-rank) เพื่อเป็นคำตอบ

2.5) ประเมินประสิทธิภาพโมเดลด้วยการคำนวณค่า Precision Recall และ MAP

ในแต่ละขั้นตอนการทำงานจะแสดงรายละเอียดดังต่อไปนี้ ในส่วนของการเรียนรู้ K-means clustering จะถูกใช้ในการสร้างรหัสของ Reference dataset สำหรับชุดข้อมูลมัลติมีเดียในแต่ละประเภท (Media) ได้แก่ รูปภาพ ข้อความ วิดีโอ เสียง และ 3 มิติ สมมติว่ามีการเข้ารหัส Reference dataset จำนวน N รายการ โดยมีความยาวเป็น c ขึ้นอยู่กับจำนวนกลุ่มข้อมูล (256 clusters, c = 8) สามารถแสดงได้ดังนี้ $B = \{b_i \in \{0,1\}^c \mid i = 1, 2, ..., N\}$ ถ้าไม่มีการสูญเสียข้อมูลในการเข้ารหัสข้อมูล 256 กลุ่ม จากจำนวน 8 บิต จะได้ว่า b_i คือ รหัสดัชนี (Index code) $x_i \in \{0, 1\}^8$ ตารางดัชนีจะมี 2^8 คือ 256 รายการ (Entries) โดยแต่ละรายการจะได้ว่า $E_X = \{b_i \mid x_i = X\}$ สำหรับแทนค่าเฉพาะของรหัสดัชนี X และบรรจุ (attaches) ข้อมูลที่เกี่ยวข้องกับของ Reference data

สามารถสอนโมเดลด้วยการเรียนรู้แบบ Nonlinear mapping ระหว่างคำถาม (Query) ของแต่ละประเภทข้อมูล (เช่น รูปภาพ) และพื้นที่ของดัชนี (index space) ของข้อมูลอีกประเภท (เช่น วิดีโอ) ผ่าน DL โมเดลที่ได้จะใช้เพื่อทำนายค่าความสัมพันธ์ (Relevance scores) ของดัชนีสำหรับคำถามใด ๆ สามารถรวบรวมข้อมูลของชุดคำถามเพื่อใช้ในการทดสอบของแต่ละประเภทข้อมูลไว้ดังนี้ $Q = \{\boldsymbol{q}_j^* \mid j = 1, 2, ..., J\}$, โดยที่ $\boldsymbol{q}_j^*$ คือลำดับของคำถาม (jth query) และค่าความสัมพันธ์ของตัวอย่างในแต่ละประเภทไว้ดังนี้ $\boldsymbol{q}_j^*$ อยู่ในรูปแบบ $\boldsymbol{b}_{jk}^* \mid k = 1, 2, ..., K\} \in B$, โดยที่ $\boldsymbol{b}_{jk}^*$ และ k คือลำดับความสัมพันธ์ของความสัมพันธ์ของข้อมูลตัวอย่างสำหรับ $\boldsymbol{q}_j^*$

การกำหนดความสัมพันธ์ตัวอย่างบนพื้นฐานของลาเบล (Class label information) สำหรับตัวอย่างความสัมพันธ์ของตัวอย่างของคำถามที่มีข้อมูลลาเบลเดียวกัน ค่าความสัมพันธ์ Relevance

score ของแต่ละรหัสดัชนี X สามารถคำนวณได้จากอัตราส่วนความสัมพันธ์ของรายการข้อมูลในดัชนีนั้น ดังสมการที่ 3.5

$$\boldsymbol{T}^{*}_{j\mathbf{X}} = \frac{|\{\boldsymbol{b}^{*}_{jk}|\boldsymbol{x}^{*}_{jk}=\mathbf{X}\}|}{|\boldsymbol{E}_{\mathbf{X}}|}, \tag{3.5}$$

โดยที่ $|\cdot|$ คือ set cardinality โดยที่คู่ของ query features และ relevance scores ถูกประมวลผล (compiled) บน training set และ *j*th query $\boldsymbol{q}^{*}_{j}$ จะสัมพันธ์กับคะแนนความสัมพันธ์ (relevance scores) ของจำนวนกลุ่มข้อมูล (clusters) ของรหัสดัชนี (index codes) $\{\boldsymbol{T}^{*}_{j\mathbf{X}}\}$

เครือข่ายประสาทที่เชื่อมต่ออย่างสมบูรณ์ (Fully-connected neural network) จะถูกใช้ในการเรียนรู้คุณสมบัติของคำถาม (Query feature) และ ความสัมพันธ์ของรหัสดัชนีของข้อมูลที่ใช้สอนโมเดล การทดลองนี้พยายามปรับใช้เครือข่ายให้มีขนาดเล็ก (Compact network) และมีผลลัพธ์ของความแม่นยำที่สูง (Highest MAP) ด้วยเช่นกันในแต่ละชุดข้อมูล ดังนั้นเราจึงใช้จำนวน Unit ของชั้นที่ซ่อนอยู่ (Hidden layers) เหมือนกันในทุกชุดข้อมูลคือ 8-16-8 โดยที่ชั้นนำเข้าข้อมูล (input layer) คือ จำนวนมิติ (Number of dimensions) ของข้อมูลคำถามดั้งเดิม (Raw query) และผลลัพธ์ (Output) คือ จำนวนของกลุ่มข้อมูล (Number of clusters) เพื่อให้มีประสิทธิภาพดีกว่าค่าพื้นฐาน (Baseline) การประมวลผล Input layer จะรับข้อมูลเข้า Feature representation (Raw query) ของ $\boldsymbol{q}^{*}_{j}$ และ Output layer ทำนายค่าของคะแนนความสัมพันธ์ของรหัสดัชนี (index codes) $\{\boldsymbol{P}^{*}_{j\mathbf{X}}\}$ บนพื้นฐานของ Cross-entropy loss ระหว่างการทำนาย (Predictions) $\{\boldsymbol{P}^{*}_{j\mathbf{X}}\}$ และ ค่าเป้าหมาย (target) $\{\boldsymbol{T}^{*}_{j\mathbf{X}}\}$ และมีการคำนวณ Error derivative เกี่ยวกับผลลัพธ์ของเซลล์ประสาทแต่ละเซลล์ (Each neuron) โดยการแพร่กระจายย้อนกลับ (Backward propagated) ไปยังแต่ละระดับชั้น (Layer) เพื่อปรับปรุงค่าน้ำหนัก (Weights) ของโครงข่ายประสาทเทียม (Neural Network)

ในส่วนของการค้นคืนข้อมูลจะใช้คำถามข้อมูลดิบสำหรับการค้นคืนข้อมูลมัลติมีเดีย (Given a raw query q) เราใช้โมเดลที่ผ่านการเรียนรู้แล้ว (Learned model) เพื่อประมาณค่า Relevance scores ของรหัสดัชนี $\{P_X\}$ โดยที่รหัสดัชนีถูกจัดอันดับเพื่อเลือกรหัสดัชนีที่มีคะแนนความเกี่ยวข้องสูงสุดไว้ระดับบนสุด $\{X_1, X_2, ..., X_R\}$ และการเชื่อมโยงของจุดข้อมูลอ้างอิงและรหัสดัชนีระดับสูงจะถูกดึงมาในชุดข้อมูลตัวเลือก (Candidate set) ดังนี้ $C = \{b_i|x_i \in X_r, r = 1, 2, ..., R\}$

ในส่วนของภาพที่ 3.6 นำเสนอการประยุกต์ใช้หลายดัชนี K-means (Multi-K-means) โดยนำจำนวนกลุ่มมาจัดแบ่งเป็น 2 ชุด (หรือสร้างรหัสใหม่ทั้ง 2 รหัส) แล้วนำแต่ละชุดไปเรียนรู้เพื่อสร้างโมเดล แล้วจึงนำโมเดลไปใช้ในการทำนายผลลัพธ์ สุดท้ายนำผลแต่ละชุดมาเลือกคู่แข่ง (candidate) โดยเอารหัสข้อมูลที่ได้จากผลลัพธ์ของทั้งสองโมเดลที่ซ้ำกัน ถ้าในผลลัพธ์ของทั้งสองส่วนไม่มีข้อข้อมูลซ้ำหรือข้อมูลไม่เพียงพอให้ยืดจากโมเดลที่ให้ผลความถูกต้องที่สูงกว่า

ในเรื่องความซับซ้อนของเวลาสำหรับการค้นหาเพื่อนบ้านที่ใกล้ที่สุด ส่วนใหญ่เกี่ยวข้องกับสองส่วนคือ 1) การคาดคะเนคะแนนความเกี่ยวข้อง และ 2) การจัดอันดับของรหัสดัชนี ในส่วนของเวลาที่ใช้ในการทำนายคะแนนความเกี่ยวข้องจะสัมพันธ์กับขนาดของโครงข่ายประสาทเทียมถือได้ว่าเป็นเวลาที่คงที่ ในส่วนของเวลาที่ใช้คือ $O(2^c \times \log 2^c = O(c \times 2^c)$ (โดยที่ c คือจำนวนบิตที่ใช้ในการสร้างตารางดัชนี) เวลาในการคำนวณเพื่อจัดเรียงรหัสดัชนีทั้งหมดตามคะแนนความเกี่ยวข้องสำหรับการจัดอันดับรหัสดัชนี วิธีการที่นำเสนอจะไม่มีการจัดอันดับผู้สมัคร (Candidate) ใหม่ (Re-ranking) หลังจากดัชนีได้รับการจัดอันดับดัชนีแล้ว ดังนั้นจึงสามารถข้ามเวลาคงที่สำหรับการคำนวณระยะทาง Hamming ได้ การคำนวณของผู้สมัครคือการคำนวณระยะทาง Hamming กับคำถามสำหรับผู้สมัครทั้งหมด ใช้เวลา $s \cdot |C|$ โดยที่ s คือเวลาคำนวณค่าคงที่เล็กน้อยของระยะ Hamming โดยปกติชุดผู้สมัคร C จะเป็นเศษส่วนของชุดข้อมูลอ้างอิง B ดังนั้นเวลาในการคำนวณจึงสามารถลดลงได้มากเมื่อเทียบกับการค้นหาแบบละเอียด ที่น่าสังเกตคือคุณภาพของชุดผู้สมัครนั้นมีค่าสูง โดยแสดงค่าความแม่นยำในการค้นหาไว้ในส่วนการทดลองในบทที่ 4

บทที่ 4
ผลการวิจัย

ในส่วนของผลการทดลอง งานวิจัยทำการทดลองเกี่ยวกับชุดข้อมูลมัลติมีเดียมาตรฐาน 4 ชุดข้อมูลที่นิยมใช้กันอย่างแพร่หลาย เพื่อตรวจสอบประสิทธิภาพของโมเดลการเรียนรู้ดัชนี K-means ทำงานบนเครื่องคอมพิวเตอร์ที่มีคุณสมบัติดังนี้ Processor 11th Gen Intel(R) Core(TM) i3-1115G4 @3.00GHz RAM 8.00 GB เมื่อเปรียบเทียบกับวิธีพื้นฐาน (baseline) แล้วผลลัพธ์ที่ได้แสดงให้เห็นถึงประสิทธิผลของวิธีการที่นำเสนอดังจะแสดงดังต่อไปนี้

4.1 ผลการค้นคืนข้อมูลด้วยตารางดัชนีแบบกลับด้านของรหัส K-means

ตารางที่ 4.1 เปรียบเทียบค่า MAP ของผลการค้นคืนข้อมูลด้วยรหัส K-means ชุดข้อมูล Wiki

Top	I -> T			T -> I		
	5-bits	8-bits	10-bits	5-bits	8-bits	10-bits
1	**0.1212**	0.1053	0.1140	**0.1472**	0.1255	0.0765
5	**0.0910**	0.0819	0.0690	0.0531	**0.0669**	0.0517
10	**0.0742**	0.0672	0.0509	**0.0466**	0.0465	0.0408
15	**0.0712**	0.0546	0.0434	**0.0440**	0.0389	0.0343
20	**0.0693**	0.0465	0.0387	**0.0390**	0.0345	0.0301
25	**0.0657**	0.0417	0.0354	**0.0357**	0.0311	0.0277
30	**0.0645**	0.0382	0.0328	**0.0326**	0.0286	0.0258
35	**0.0638**	0.0355	0.0305	**0.0303**	0.0269	0.0245
40	**0.0631**	0.0335	0.0287	**0.0289**	0.0254	0.0235
45	**0.0623**	0.0320	0.0271	**0.0270**	0.0244	0.0226
50	**0.0609**	0.0307	0.0259	**0.0258**	0.0234	0.0218

ตารางที่ 4.2 เปรียบเทียบค่า MAP ของผลการค้นคืนข้อมูลด้วยรหัส K-means ชุดข้อมูล NUS-WIDE

Top	I -> T			T -> I		
	5-bits	8-bits	9-bits	5-bits	8-bits	9-bits
1	**0.3290**	0.2937	0.2905	0.2760	**0.2792**	0.2637
5	**0.2444**	0.2306	0.2406	0.1888	0.1710	**0.1913**
10	0.2113	0.2179	**0.2221**	0.1657	0.1512	**0.1636**
15	0.1988	**0.2123**	0.2115	0.1557	0.1452	**0.1540**
20	0.1922	**0.2077**	0.2065	0.1454	0.1400	**0.1487**
25	0.1892	**0.2053**	0.2019	0.1407	0.1342	**0.1434**
30	0.1865	**0.2027**	0.1995	0.1382	0.1321	**0.1397**
35	0.1827	**0.2002**	0.1975	0.1354	0.1281	**0.1382**
40	0.1789	**0.1986**	0.1957	0.1326	0.1261	**0.1365**
45	0.1774	**0.1969**	0.1943	0.1306	0.1248	**0.1348**
50	0.1769	**0.1952**	0.1929	0.1289	0.1232	**0.1332**

จากตารางที่ 4.1 และ 4.2 แสดงการเปรียบเทียบค่า MAP ของ 50 อันดับแรกของผลการค้นคืนข้อความโดยใช้คำค้นด้วยรูปภาพ (I -> T) และ การค้นคืนรูปภาพด้วยข้อความ (T-> I) ด้วยรหัส K-means ตามจำนวนบิตต่าง ๆ สำหรับชุดข้อมูล Wiki และ NUS-WIDE ตามลำดับ โดยตัวเลขหนาคือค่าที่ดีที่สุด แสดงให้เห็นว่าผลลัพธ์ความถูกต้องไม่คงที่ กล่าวคือเมื่อเพิ่มจำนวนนวนบิตก็ไม่ได้ผลลัพธ์ที่ดีขึ้นเสมอไป แสดงให้เห็นถึงรหัสจากการจัดกลุ่ม K-means ยังมีประสิทธิภาพน้อย

4.2 ประสิทธิภาพการค้นคืนข้อมูลข้ามมีเดีย

ตารางที่ 4.3 การเปรียบเทียบค่า Precision ของ 100 อันดับแรกของผลการค้นคืนข้อความโดยใช้คำค้นด้วยรูปภาพ (I -> T) ด้วยรหัสไบนารี่ 8 และ 10 บิต สำหรับชุดข้อมูล Wiki

Top	k-means 8-bits	k-means 10-bits	Our 8-bits	Our 10-bits
1	0.1053	0.1140	0.2165	**0.2323**
5	0.1036	0.1085	0.2118	**0.2317**
10	0.1049	0.1089	0.2137	**0.2296**
20	0.1038	0.1118	0.2141	**0.2291**
50	0.1026	0.1096	0.2153	**0.2235**
100	0.1043	0.1106	0.1527	**0.1595**

ตารางที่ 4.4 การเปรียบเทียบค่า Recall ของ 100 อันดับแรกของผลการค้นคืนข้อความโดยใช้คำค้นด้วยรูปภาพ (I -> T) ด้วยรหัสไบนารี่ 8 และ 10 บิต สำหรับชุดข้อมูล Wiki

Top	k-means 8-bits	k-means 10-bits	Our 8-bits	Our 10-bits
1	0.0004	0.0005	0.0009	**0.0010**
5	0.0022	0.0024	0.0044	**0.0048**
10	0.0044	0.0047	0.0090	**0.0095**
20	0.0087	0.0095	0.0180	**0.0190**
50	0.0216	0.0234	0.0453	**0.0461**
100	0.0439	0.0469	0.0698	**0.0713**

ตารางที่ 4.5 การเปรียบเทียบค่า Precision ของ 100 อันดับแรกของผลการค้นคืนรูปโดยใช้คำค้นด้วยข้อความ (T -> I) ด้วยรหัสไบนารี่ 8 และ 10 บิต สำหรับชุดข้อมูล Wiki

Top	k-means 8-bits	k-means 10-bits	Our 8-bits	Our 10-bits
1	0.1255	0.0765	0.6032	**0.6335**
5	0.1108	0.1027	0.5224	**0.6381**
10	0.1092	0.1137	0.4967	**0.6384**
20	0.1132	0.1092	0.4620	**0.6334**
50	0.1116	0.1124	0.3990	**0.6152**
100	0.1094	0.1108	0.2507	**0.3511**

ตารางที่ 4.6 การเปรียบเทียบค่า Recall ของ 100 อันดับแรกของผลการค้นคืนรูปโดยใช้คำค้นด้วยข้อความ (T -> I) ด้วยรหัสไบนารี่ 8 และ 10 บิต สำหรับชุดข้อมูล Wiki

Top	k-means 8-bits	k-means 10-bits	Our 8-bits	Our 10-bits
1	0.0005	0.0003	0.0026	**0.0028**
5	0.0024	0.0022	0.0114	**0.0139**
10	0.0047	0.0047	0.0213	**0.0278**
20	0.0097	0.0092	0.0393	**0.0551**
50	0.0239	0.0238	0.0850	**0.1319**
100	0.0466	0.0466	0.1123	**0.1561**

ตารางที่ 4.7 การเปรียบเทียบค่า Precision และ Recall ของ 100 อันดับแรกของผลการค้นคืนรูปโดยใช้คำค้นด้วยข้อความ (I -> T) ด้วยรหัสไบนารี่ 8 บิต สำหรับชุดข้อมูล MIRFlickr

Top	k-means precision	k-means recall	Our precision	Our recall
1	0.5801	0.0001	**0.8266**	**0.0001**
5	0.5619	0.0003	**0.8050**	**0.0005**
10	0.5556	0.0006	**0.8038**	**0.0009**
20	0.5508	0.0012	**0.7923**	**0.0040**
50	0.5456	0.0031	**0.7830**	**0.0045**
100	0.5463	0.0061	**0.7692**	**0.0086**

ตารางที่ 4.8 การเปรียบเทียบค่า Precision และ Recall ของ 100 อันดับแรกของผลการค้นคืนรูปโดยใช้คำค้นด้วยข้อความ (T -> I) ด้วยรหัสไบนารี่ 8 บิต สำหรับชุดข้อมูล MIRFlickr

Top	k-means precision	k-means recall	Our precision	Our recall
1	0.5658	0.0001	**0.7129**	**0.0001**
5	0.5821	0.0003	**0.7739**	**0.0004**
10	0.5713	0.0010	**0.7789**	**0.0013**
20	0.5692	0.0013	**0.7672**	**0.0018**
50	0.5583	0.0031	**0.7631**	**0.0044**
100	0.5567	0.0063	**0.7535**	**0.0085**

จากตารางที่ 4.3-4.6 แสดงผลการทดลองการค้นคืนข้อมูลข้ามมีเดีย Precision และ Recall โดยการค้นหารูปด้วยคำค้นข้อความ (T -> I) และค้นหาข้อความด้วยคำค้นรูป (I -> T) บนชุดข้อมูล Wiki ด้วยข้อมูล 8 บิต และ 10 บิต ซึ่งผลลัพธ์แสดงให้เห็นว่าการใช้กลุ่มข้อมูลจำนวนมากกว่าส่งผลให้ความถูกต้องของการค้นคืนข้อมูลที่สูงขึ้น เนื่องจากจำนวนบิตหรือกลุ่มข้อมูลที่สูงขึ้นสามารถจำแนกความแตกต่างข้อมูลได้มากขึ้น สำหรับตารางที่ 4.7-4.8 แสดงผลการค้นคืนของชุดข้อมูล MIRFlickr 8 บิต ซึ่งผลลัพธ์แสดงให้เห็นว่าวิธีการที่นำเสนอมีความถูกต้องมากกว่าตารางแบบดั้งเดิม

ตารางที่ 4.9 การเปรียบเทียบค่า MAP ของ 50 อันดับแรกของผลการค้นคืนข้อความโดยใช้คำค้นด้วยรูปภาพ (I -> T) ด้วยรหัสไบนารี่ 5 บิต 8 บิต และ 10 บิต สำหรับชุดข้อมูล Wiki

Top	k-means 5-bits	k-means 8-bits	k-means 10-bits	Our 5-bits	Our 8-bits	Our 10-bits
1	0.1212	0.1053	0.1140	0.2092	**0.2323**	0.2165
5	0.0910	0.0819	0.0690	0.1989	**0.2140**	0.2055
10	0.0742	0.0672	0.0509	0.1880	**0.2040**	0.2029
15	0.0712	0.0546	0.0434	0.1797	0.2009	**0.2010**
20	0.0693	0.0465	0.0387	0.1745	0.1976	**0.1991**
25	0.0657	0.0417	0.0354	0.1722	0.1929	**0.1972**
30	0.0645	0.0382	0.0328	0.1697	0.1895	**0.1961**
35	0.0638	0.0355	0.0305	0.1684	0.1866	**0.1949**
40	0.0631	0.0335	0.0287	0.1666	0.1835	**0.1941**
45	0.0623	0.0320	0.0271	0.1664	0.1807	**0.1927**
50	0.0609	0.0307	0.0259	0.1659	0.1785	**0.1918**

ตารางที่ 4.10 การเปรียบเทียบค่า MAP ของ 50 อันดับแรกของผลการค้นคืนรูปโดยใช้คำค้นด้วยข้อความ (T -> I) ด้วยรหัสไบนารี่ 5 บิต 8 บิต และ 10 บิต สำหรับชุดข้อมูล Wiki

Top	k-means 5-bits	k-means 8-bits	k-means 10-bits	Our 5-bits	Our 8-bits	Our 10-bits
1	0.1472	0.1255	0.0765	0.3781	0.6032	**0.6595**
5	0.0531	0.0669	0.0517	0.2262	0.4501	**0.6184**
10	0.0466	0.0465	0.0408	0.1779	0.3853	**0.6080**
15	0.0440	0.0389	0.0343	0.1453	0.3495	**0.6004**
20	0.0390	0.0345	0.0301	0.1280	0.3248	**0.5961**
25	0.0357	0.0311	0.0277	0.1160	0.3033	**0.5903**
30	0.0326	0.0286	0.0258	0.1072	0.2835	**0.5850**
35	0.0303	0.0269	0.0245	0.1003	0.2710	**0.5797**
40	0.0289	0.0254	0.0235	0.0948	0.2592	**0.5737**
45	0.0270	0.0244	0.0226	0.0912	0.2498	**0.5678**
50	0.0258	0.0234	0.0218	0.0862	0.2416	**0.5603**

ตารางที่ 4.11 การเปรียบเทียบค่า MAP ของ 50 อันดับแรกของผลการค้นคืนข้ามมีเดีย ด้วยรหัสไบนารี่ 8 บิต สำหรับชุดข้อมูล Xmedia

Top	k-means (I -> T)	Our (I -> T)	k-means (T -> I)	Our (T -> I)
1	0.0450	**0.2340**	0.0380	**0.2100**
5	0.0261	**0.1598**	0.0219	**0.1512**
10	0.0188	**0.1368**	0.0167	**0.1312**
15	0.0152	**0.1246**	0.0141	**0.1229**
20	0.0131	**0.1189**	0.0123	**0.1176**
25	0.0115	**0.1134**	0.0111	**0.1129**
30	0.0103	**0.1093**	0.0101	**0.1082**
35	0.0094	**0.1050**	0.0093	**0.1044**
40	0.0087	**0.1019**	0.0086	**0.1007**
45	0.0082	**0.0977**	0.0081	**0.0968**
50	0.0078	**0.0946**	0.0077	**0.0939**

ตารางที่ 4.12 การเปรียบเทียบค่า MAP ของ 50 อันดับแรกของผลการค้นคืนข้ามมีเดีย ด้วยรหัสไบนารี่ 8 บิต สำหรับชุดข้อมูล MIRFlickr

Top	k-means (I -> T)	Our (I -> T)	k-means (T -> I)	Our (T -> I)
1	0.5801	**0.8266**	0.5658	**0.7129**
5	0.4817	**0.7739**	0.4721	**0.7009**
10	0.4440	**0.7603**	0.4420	**0.6958**
15	0.4273	**0.7452**	0.4255	**0.6910**
20	0.4160	**0.7350**	0.4129	**0.6772**
25	0.4081	**0.7269**	0.4029	**0.6733**
30	0.4013	**0.7213**	0.3974	**0.6728**
35	0.3958	**0.7155**	0.3925	**0.6670**
40	0.3910	**0.7144**	0.3888	**0.6595**
45	0.3875	**0.7118**	0.3861	**0.6575**
50	0.3849	**0.7096**	0.3816	**0.6573**

ตารางที่ 4.13 การเปรียบเทียบค่า MAP ของ 50 อันดับแรกของผลการค้นคืนข้ามมีเดีย ด้วยรหัสไบนารี่ 8 บิต สำหรับชุดข้อมูล NUS-WIDE

Top	k-means (I -> T)	Our (I -> T)	k-means (T -> I)	Our (T -> I)
1	0.2937	**0.4480**	0.2792	**0.5761**
5	0.2306	**0.3395**	0.1710	**0.4960**
10	0.2179	**0.3255**	0.1512	**0.4682**
15	0.2123	**0.3186**	0.1452	**0.4574**
20	0.2077	**0.3159**	0.1400	**0.4521**
25	0.2053	**0.3140**	0.1342	**0.4251**
30	0.2027	**0.3128**	0.1321	**0.4086**
35	0.2002	**0.3114**	0.1281	**0.4003**
40	0.1986	**0.3097**	0.1261	**0.3949**
45	0.1969	**0.2957**	0.1248	**0.3850**
50	0.1952	**0.2865**	0.1232	**0.3809**

ตารางที่ 4.14 การเปรียบเทียบค่า MAP ของ 50 อันดับแรกของผลการค้นคืนข้ามมีเดีย ด้วยรหัสไบนารี่ 9 บิต สำหรับชุดข้อมูล NUS-WIDE

Top	k-means (I -> T)	Our (I -> T)	k-means (T -> I)	Our (T -> I)
1	0.2905	**0.7010**	0.2637	**0.5970**
5	0.2406	**0.6250**	0.1913	**0.5562**
10	0.2221	**0.6039**	0.1636	**0.5332**
15	0.2115	**0.5868**	0.1540	**0.4908**
20	0.2065	**0.5700**	0.1487	**0.4908**
25	0.2019	**0.5670**	0.1434	**0.4554**
30	0.1995	**0.5624**	0.1397	**0.4461**
35	0.1975	**0.5601**	0.1382	**0.4369**
40	0.1957	**0.5575**	0.1365	**0.4300**
45	0.1943	**0.5567**	0.1348	**0.4268**
50	0.1929	**0.5550**	0.1332	**0.4225**

ตารางที่ 4.9-4.10 เปรียบเทียบค่า MAP ของโมเดล โดยการค้นหารูปด้วยข้อความ (T -> I) และค้นหาข้อความด้วยคำค้นรูป (I -> T) บนชุดข้อมูล Wiki ด้วยข้อมูล 5 บิต 8 บิต และ 10 บิต โดยในภาพรวมการแทนข้อมูลด้วย 10 บิตให้ผลลัพธ์ความถูกต้องที่สูงกว่า 5 และ 8 บิต สำหรับตารางที่ 4.11-4.13 แสดงผลการค้นคืนของชุดข้อมูล Xmedia MIRFlickr และ NUS-WIDE จำนวน 8 บิต ตามลำดับ และตารางที่ 4.14 แสดงผลการค้นคืนของชุดข้อมูล NUS-WIDE จำนวน 9 บิต ซึ่งผลลัพธ์ความถูกต้องของแนวคิดที่นำเสนอสูงกว่าการใช้ดัชนี K-means แบบดั้งเดิม และเมื่อเพิ่มจำนวนบิตผลลัพธ์ของวิธีที่นำเสนอให้ผลลัพธ์ที่ถูกต้องเพิ่มขึ้นในอัตราส่วนที่สูงกว่าแบบเดิม

กล่าวโดยสรุป สำหรับการเปรียบเทียบผลการค้นคืนข้อมูลด้วยค่า Precision Recall และ MAP แสดงให้เห็นว่าโมเดลที่นำเสนอมีประสิทธิภาพสูงกว่า Baseline และเมื่อใช้จำนวนกลุ่มข้อมูลหรือจำนวนเลขบิตที่สูงขึ้นส่งผลให้ประสิทธิภาพการค้นคืนสูงขึ้นตามลำดับ

ตารางที่ 4.15 แสดงค่า MAP ของ 1 รายการแรกของผลการค้นคืนแบบ Multi-K-means 5 บิต ของชุดข้อมูล Wiki Xmedia และ MIRFlickr

	k-means 5-bits	Our 5-bits	Our Multi-K-means 5 bits
Wiki			
I -> T	0.1212	0.2092	**0.2482**
T -> I	0.1472	0.2713	**0.3362**
Xmedia			
I -> T	0.0500	0.059	**0.117**
T -> I	0.0402	0.052	**0.072**
MIRFlickr			
I -> T	0.5550	0.5634	**0.5861**
T -> I	0.5335	0.5778	**0.5801**

ในส่วนตารางที่ 15 แสดงค่า MAP บนชุดข้อมูล Wiki Xmedia และ MIRFlickr ของ 1 รายการแรกจากผลการค้นคืนแบบ Multi-K-means ด้วยรหัส 5 บิต โดยการเอาผลโหวตรหัสข้อมูลจากจำนวน 100 ข้อมูลแรกของทั้งสองชุดข้อมูลและเลือกรายการที่ปรากฏซ้ำทั้ง 2 ชุด คือ จาก 5 บิตแรก และ 5 บิตหลังเป็นตัวเลือก ซึ่งผลลัพธ์แสดงว่าการรวมบิตโหวตทำให้ผลลัพธ์สูงขึ้นเล็กน้อย

ตารางที่ 4.16 เปรียบเทียบเวลาที่ใช้ในการค้นคืนข้อมูลหน่วยเป็นวินาที ผลลัพธ์ 50 อันดับแรก

Dataset	Reference set	Exhaustive search	Our	Difference time
WIKI				
I -> T	2866	0.1470	**0.0799**	0.0671
T -> I	2866	0.1500	**0.0800**	0.0700
Xmedia				
I -> T	5000	0.4064	**0.2721**	0.1343
T -> I	5000	0.4376	**0.2893**	0.1483
MIRFlickr				
I -> T	15902	1.5061	**0.2397**	1.2664
T -> I	15902	1.4831	**0.3341**	1.1490
NUS-WIDE				
I -> T	184711	38.4014	**0.3663**	38.0351
T -> I	184711	40.4183	**0.4010**	40.0173

จากตารางที่ 4.16 แสดงผลการเปรียบเทียบเวลาที่ใช้ (Time complexity) ในการค้นคืนข้อมูลในหน่วยวินาทีของผลลัพธ์ 50 อันดับแรก โดยใช้ข้อมูล 256 กลุ่มข้อมูล รหัส 8 บิต สำหรับ 4 ชุดข้อมูล คือ Wiki Xmedia MIRFlickr และ NUS-WIDE (ตัวอักษรหนาหมายถึงผลลัพธ์ที่ดีกว่า) จากผลการทดลองสังเกตุได้ว่าวิธีการที่นำเสนอใช้เวลาในการประมวลผลน้อยกว่า และผลต่างของเวลาประมวลผลที่สูงขึ้นเมื่อขนาดชุดข้อมูลที่ใหญ่ขึ้น แสดงให้เห็นว่าวิธีการที่นำเสนอเหมาะสมกับการประมวลผลกับชุดข้อมูลที่มีขนาดใหญ่

บทที่ 5
สรุป อภิปรายผล และข้อเสนอแนะ

งานวิจัยนี้นำเสนอกระบวนการค้นคืนข้อมูลแบบใหม่โดยการประยุกต์ใช้ค่าความสัมพันธ์ของตารางดัชนีแบบกลับด้านของรหัส K-means ของ 4 ชุดข้อมูล และการประยุกต์ใช้ Multi-K-means ซึ่งสามารถสรุป อภิปรายผล และข้อเสนอแนะ ได้ดังนี้

5.1 สรุป และอภิปลายผลการวิจัย

1) ตารางดัชนีที่สร้างจากรหัสการจัดกลุ่ม K-means ให้ผลลัพธ์ค่าความถูกต้องที่ต่ำ เมื่อเพิ่มจำนวนกลุ่มข้อมูลหรือจำนวนบิตก็จะยังมีค่าความถูกต้องที่ไม่สูงขึ้นมาก และค่าผลลัพธ์ก็ไม่คงที่เสมอไป แสดงว่าจำนวนบิตที่สูงขึ้นอาจจะไม่สอดคล้องกับค่าความถูกต้อง อีกทั้งเมื่อเพิ่มจำนวนบิตส่งผลให้ต้องใช้หน่วยความจำและเวลาในการประมวลผลก็จะสูงขึ้นตามลำดับ

2) เมื่อนำรหัสกลุ่ม K-means มาเรียนรู้ความสัมพันธ์ระหว่างค่าความสัมพันธ์ที่ประยุกต์ใช้จากค่าความน่าจะเป็นของข้อมูลที่ถูกต้องของแต่ละกลุ่มกับข้อมูลคำถามที่เป็นข้อมูลดิบผ่านการเรียนรู้เชิงลึกของวิธีการที่นำเสนอ ผลลัพธ์แสดงให้เห็นว่าโมเดลสามารถเพิ่มความถูกต้องแม่นยำด้วยค่า Precision Recall และ MAP เมื่อเปรียบเทียบกับการใช้รหัส K-means แบบเดิม จากผลลัพธ์ดังกล่าวสามารถอภิปรายผลได้ว่าโมเดลสามารถเรียนรู้คุณลักษณะของชุดข้อมูลคำถามที่เป็นข้อมูลดิบที่เชื่อมโยงกับค่าความสัมพันธ์ของความน่าจะเป็นของค่าความถูกต้องในแต่ละกลุ่มได้อย่างมีประสิทธิภาพ อีกทั้งยังสามารถลดเวลาที่ใช้ในการค้นคืนข้อมูลได้ เนื่องจากไม่จำเป็นต้องเปรียบเทียบข้อมูลระหว่างคำถามกับทุกกลุ่มข้อมูลในตารางดัชนี อีกทั้งจำนวนข้อมูลในกลุ่มที่มีค่าความน่าจะเป็นสูงจะมีจำนวนข้อมูลอื่น ๆ ที่ไม่เกี่ยวข้องน้อย และเมื่อได้ข้อมูลจากกลุ่มที่มีความน่าจะเป็นสูงสุดแล้วโมเดลใหม่จำเป็นต้องมีการจัดเรียงข้อมูลอีกครั้งเพื่อลดเวลาในส่วนนี้ด้วย สำหรับโมเดลที่นำเสนอเมื่อใช้จำนวนกลุ่มข้อมูลหรือจำนวนเลขบิตที่สูงขึ้นส่งผลให้ประสิทธิภาพการค้นคืนสูงขึ้นตามลำดับ แต่หน่วยความจำและเวลาที่ใช้ก็จะสูงขึ้นตามขนาดของจำนวนบิตขึ้น ผู้ที่สนใจสามารถปรับเปลี่ยนเครื่องคอมพิวเตอร์ให้เหมาะสมกับประสิทธิภาพที่ต้องการ ดังนั้นจึงควรเลือกใช้จำนวนบิตและขนาดของเครื่องคอมพิวเตอร์ที่เหมาะสมกับข้อมูลแต่ละชุดข้อมูล อีกทั้งสามารถประยุกต์ใช้การค้นคืนข้อมูลที่ใช้หลายรหัส K-means โดยวิธีการนี้สามารถเพิ่มความถูกต้องได้ ซึ่งสอดคล้องการงานวิจัยของ Zhang และคณะ [34] ที่ใช้หลาย Hash code ในการค้นคืนข้อมูลข้ามมีเดียเพื่อเพิ่มประสิทธิภาพของวิธีการเดิม จากผลการทดลองแสดงให้เห็นว่าการทดลองนี้เป็นไปตามสมมติฐานที่ตั้งไว้ คือ การเรียนรู้เชิงลึกเพื่อหาความสัมพันธ์ระหว่างคำค้นข้อมูลดิบกับค่าความสัมพันธ์ของแต่ละกลุ่มข้อมูล สามารถเพิ่มประสิทธิภาพการค้นคืนข้อมูลแบบข้ามมีเดียผ่านตารางดัชนีกลับด้านของ K-means เดิมได้

5.2 ข้อเสนอแนะ

1) ผู้ที่สนใจสามารถทดลองประยุกต์ใช้ Hash code จากวิธีการอื่นๆ มาใช้ในการสร้างดัชนีและคำนวณค่าความสัมพันธ์และทดสอบความถูกต้อง

2) ผู้ที่สนใจสามารถทดลองใช้จำนวนกลุ่มหรือบิตที่สูงขึ้นในการสร้างตารางดัชนีและการเรียนรู้ โดยเลือกใช้เครื่องคอมพิวเตอร์ที่มีสมรรถนะเหมาะสมกับขนาดข้อมูล หรือใช้เครื่องคอมพิวเตอร์ที่มีขนาดที่สูงขึ้นและจำนวนบิตที่มากกว่าการทดลองครั้งนี้

3) ผู้ที่สนใจสามารถทดลองประยุกต์ใช้หลาย K-means ที่รหัสยาวขึ้น หรือมากกว่า 2 รหัสขึ้นไป เพื่อดูประสิทธิภาพ

4) ข้อสังเกตจากการทดลอง ถ้าข้อมูลที่ใช้ในการเรียนรู้น้อยเกินไป จะส่งผลให้ประสิทธิภาพที่ได้มีค่าน้อยด้วย

บรรณานุกรม

[1] Y. Peng, X. Huang, and Y. Zhao, "An overview of cross-media retrieval: concepts, methodologies, benchmarks, and challenges," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 28, No. 9, pp. 2372-85, 2018.

[2] Z. B. Wu and J. Q. Yu, "Vector quantization: a review," *Frontiers of Information Technology and Electronic Engineering*, Vol. 20, No. 4, pp. 507-524, 2019.

[3] S. Ercoli, S., M. Bertini, M., and A. Del Bimbo, "Compact hash codes for efficient visual descriptors retrieval in large scale databases," *IEEE Transactions on Multimedia*, Vol. 19, No. 11, pp. 2521-2532, 2017.

[4] J. Yadav and M. Sharma, "A Review of K-mean Algorithm," *International journal of engineering trends and technology*, Vol. 4, No. 7, pp. 2972-2976, 2013.

[5] H. J´egou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 1, pp. 117-128, 2011.

[6] N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G. R. Lanckriet, R. Levy, and N. Vasconcelos, "A new approach to cross-modal multimedia retrieval," *ACM international conference on Multimedia (ACM-MM)*, pp. 251–260, 2010.

[7] M. J. Huiskes and M. S. Lew, "The mir flickr retrieval evaluation," in *Proceedings of the ACM International Conference on Multimedia Information Retrieval*, pp. 39-43, 2008.

[8] T. S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, "Nus-wide: a real-world web image database from national university of Singapore," in *Proceedings of the ACM International Conference on Image and Video Retrieval*, pp. 48, 2009.

[9] Y. Peng, J. Qi and Y. Yuan, "Modality-specific Cross-modal Similarity Measurement with Recurrent Attention Network," *IEEE Transactions on Image Processing (TIP)*, Vol. 27, No. 11, pp. 5585-5599, Nov. 2018.

[10] J. Wang, H. T. Shen, J. Song, and J. Ji, “Hashing for similarity search: a survey,” *ArXiv:1408.2927*, 2014.

[11] K-means clustering. [online]. Resource https://towardsdatascience.com/ (April 30, 2022)

[12] Y. Arkeman, N. A. Wahanani, and A. Kustiyo, “Clustering K-Means Optimization with Multi-Objective Genetic Algorithm,” *International Journal of Electrical and Computer Sciences IJECS-IJENS*, Vol. 12, No. 5, pp. 61-66, 2012.

[13] J. Selvakumar, A. Lakshmi, and T. Arivoli, “Brain tumor segmentation and its area calculation in brain MR images using K-mean clustering and Fuzzy C-mean algorithm,” in *IEEE-international conference on advances in engineering, science and management (ICAESM-2012)*, pp. 186-190, Mar. 2012.

[14] P. Shan, “Image segmentation method based on K-mean algorithm,” *EURASIP Journal on Image and Video Processing*, Vol. 1, No. 81, 2018.

[15] M. Y. Kamil and A. M. Salih, “Mammography Images Segmentation via Fuzzy C-mean and K-mean,” *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 1, pp. 22-29, 2019.

[16] P. T. T. Binh, T. N. Le, and N. P. Xuan, “Advanced SOM and K Mean Method for Load Curve Clustering,” *International Journal of Electrical and Computer Engineering*, Vol. 8, No. 6, pp. 4829, 2018.

[17] W. Dong, M. Charikar, and K. Li, “Asymmetric distance estimation with sketches for similarity search in high-dimensional spaces,” in *Proceedings of ACM International Conference on Information Retrieval (SIGIR)*, pp. 128-130, 2008.

[18] A. Gordo, F. Perronnin, Y. Gong, and S. Lazebnik, “Asymmetric distances for binary embeddings,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 1, pp. 33-47, 2014.

[19] J. Wang, H. T. Shen, S. Yan, N. Yu, S. Li, and J. Wang, "Optimized distances for binary code ranking," in *Proceedings of ACM International conference on Multimedia (ACMMM)*, pp. 517-526, 2014

[20] D. Sheel, "Deep learning," *Revista ABB*, 1, pp. 70-71, 2019.

[21] C. Y. Chiu, J. S. Chiu, S. Markchit, and S. H. Chou, "Effective product quantization-based indexing for nearest neighbor search," *Multimedia Tools and Applications*, Vol. 78, No. 3, 2877-2895, 2019.

[22] J. Song, H. T. Shen, J. Wang, Z. Huang, N. Sebe, and J. Wang, "A distance-computation-free search scheme for binary code databases," *IEEE Transactions on Multimedia*, Vol. 18, No. 3, pp. 484-495, 2016.

[23] C. Y. Chiu and S. Markchit, "Effective and efficient indexing in cross-modal hashing-based datasets," *Signal Processing: Image Communication*, Vol. 80, pp. 115650, 2020.

[24] S. Zhang, C. Zhang, Z. You, R. Zheng and B. Xu, "Asynchronous stochastic gradient descent for DNN training," in *2013 IEEE International Conference on Acoustics*, Speech and Signal Processing, pp. 6660-6663. IEEE, 2013.

[25] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey and S. Khudanpur, "X-vectors: Robust dnn embeddings for speaker recognition," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5329-5333. IEEE, 2018.

[26] J. Zhang, Y. Zheng, D. Qi, R. Li and X. Yi, "DNN-based prediction model for spatio-temporal data," in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 1-4, 2016.

[27] M. Juuti, S. Szyller, S. Marchal and N. Asokan, "PRADA: protecting against DNN model stealing attacks," in *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*, pp. 512-527. IEEE, 2019.

[28] Y. Chen, J. He, X. Zhang, C. Hao and D. Chen, “Cloud-DNN: An open framework for mapping DNN models to cloud FPGAs,” in *Proceedings of the 2019 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays,* pp. 73-82, 2019.

[29] X. Yang, F. Li and H. Liu, “A survey of DNN methods for blind image quality assessment,” *IEEE Access*, 7, 123788-123806, 2019.

[30] D. Narayanan, A. Harlap, A. Phanishayee, V. Seshadri, N. R. Devanur, G. R. Ganger and M. Zaharia, “PipeDream: generalized pipeline parallelism for DNN training”. in *Proceedings of the 27th ACM Symposium on Operating Systems Principles*, pp. 1-15, 2019.

[31] X. Li, W. Wu, Y. H. Yuan, S. Pan, and X. Shen, “Online unsupervised cross-view discrete hashing for large-scale retrieval,” *Applied Intelligence*, 1-13, 2022.

[32] S. Sharma, V. Gupta, and M. Juneja, “A novel unsupervised multiple feature hashing for image retrieval and indexing (MFHIRI),” *Journal of Visual Communication and Image Representation*, 103467, 2022.

[33] D. Zhang, X. J. Wu, H. F. Yin, & J. Kittler, “MOON: Multi-hash codes joint learning for cross-media retrieval.” Pattern Recognition Letters, 151, 19-25, 2021.

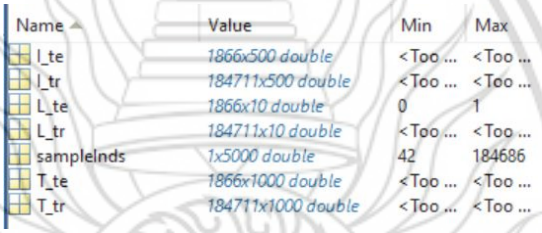
[34] B. Juba, & H. S. Le, “Precision-recall versus accuracy and the role of large data sets,” In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33, No. 01, pp. 4039-4048, 2019.

ภาคผนวก

ภาคผนวก ก

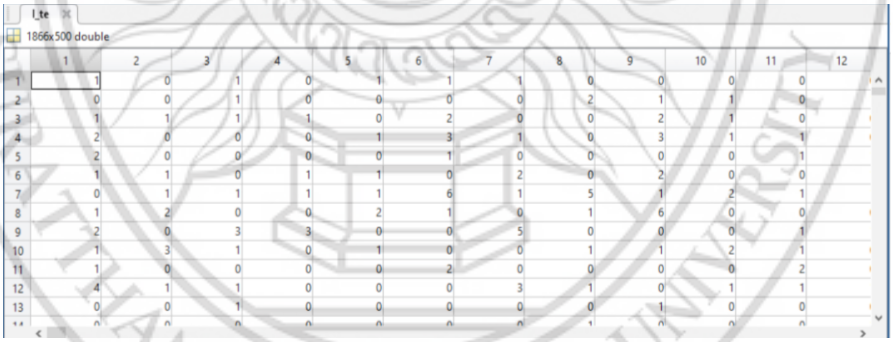
แสดงตัวอย่างรายละเอียดข้อมูลของแต่ละชุดข้อมูลดังนี้

1. ชุดข้อมูล NUS-WIDE ประกอบด้วยข้อมูล 2 ประเภทคือ ข้อความและรูปภาพ โดยข้อมูลถูกแบ่งออกเป็นข้อมูลที่ใช้ทดสอบ 1866 รายการ ที่เหลือเป็น reference set ข้อมูลรูปมีจำนวน 500 มิติ ข้อความมีจำนวน 1000 มิติ ข้อมูลลาเบลจำนวน 10 มิติ รายละเอียดต่างๆ แสดงดังภาพผนวกที่ 1-3



Name	Value	Min	Max
I_te	1866x500 double	<Too ...	<Too ...
I_tr	184711x500 double	<Too ...	<Too ...
L_te	1866x10 double	0	1
L_tr	184711x10 double	<Too ...	<Too ...
sampleInds	1x5000 double	42	184686
T_te	1866x1000 double	<Too ...	<Too ...
T_tr	184711x1000 double	<Too ...	<Too ...

ภาพผนวกที่ 1 จำนวนข้อมูลของชุดข้อมูล NUS-WIDE

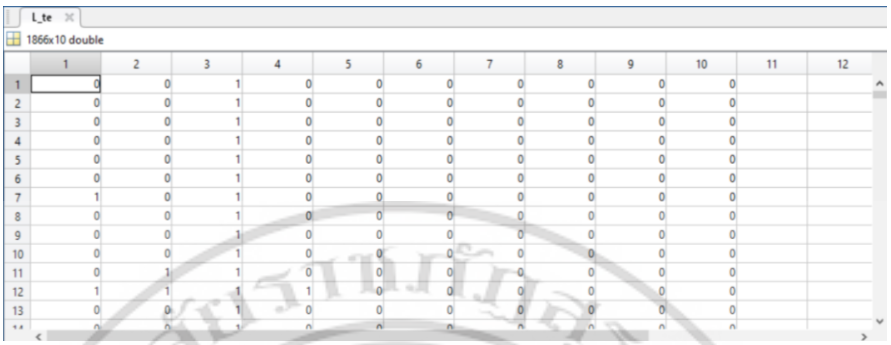


I_te

1866x500 double

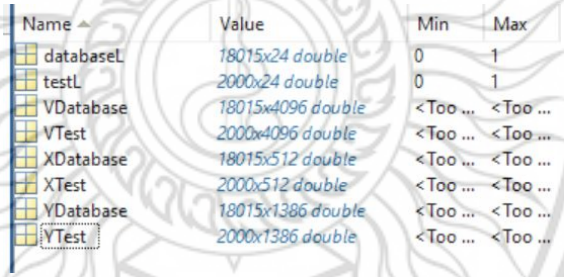
	1	2	3	4	5	6	7	8	9	10	11	12
1	1	0	1	0	1	1	1	0	0	0	0	
2	0	0	1	0	0	0	0	2	1	1	0	
3	1	1	1	1	0	2	0	0	2	1	0	
4	2	0	0	0	1	3	1	0	3	1	1	
5	2	0	0	0	0	1	0	0	0	0	1	
6	1	1	0	1	1	0	2	0	2	0	0	
7	0	1	1	1	1	6	1	5	1	2	1	
8	1	2	0	0	2	1	0	1	6	0	0	
9	2	0	3	3	0	0	5	0	0	0	1	
10	1	3	1	0	1	0	0	1	1	2	1	
11	1	0	0	0	0	2	0	0	0	0	2	
12	4	1	1	0	0	0	3	1	0	1	1	
13	0	0	1	0	0	0	0	0	1	0	0	

ภาพผนวกที่ 2 ตัวอย่างรายละเอียดข้อมูลรูปภาพของชุดข้อมูล NUS-WIDE

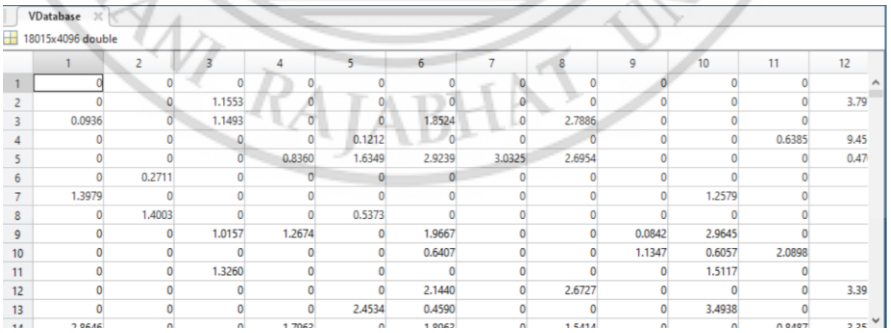


ภาพผนวกที่ 3 ตัวอย่างรายละเอียดข้อมูลข้อความของชุดข้อมูล NUS-WIDE

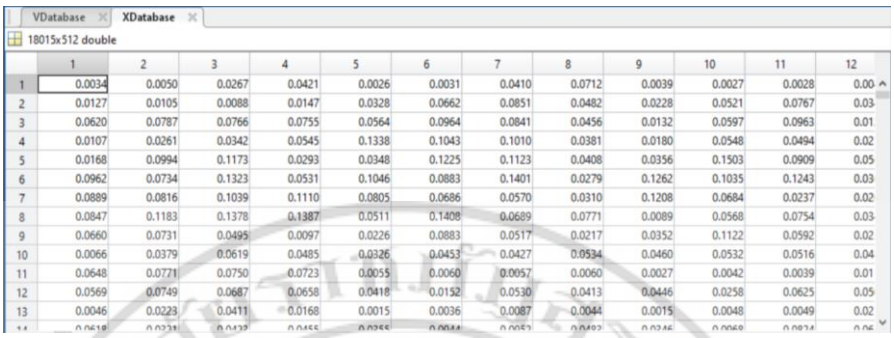
2. ชุดข้อมูล MIRflickr ประกอบด้วยข้อมูล 2 ประเภทคือ ข้อความและรูปภาพ โดยข้อมูลถูกแบ่งออกเป็นข้อมูลที่ใช้ทดสอบ 2000 รายการ ที่เหลือเป็น reference set ข้อมูลรูปภาพมีจำนวน 512 มิติ ข้อความมีจำนวน 4096 มิติ ข้อมูลลาเบลจำนวน 24 มิติ รายละเอียดต่างๆ แสดงดังภาพผนวกที่ 4-6



ภาพผนวกที่ 4 จำนวนข้อมูลของชุดข้อมูล MIRflickr

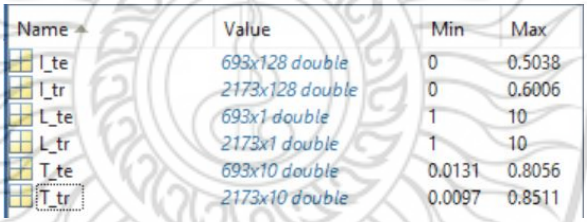


ภาพผนวกที่ 5 ตัวอย่างรายละเอียดข้อมูลรูปภาพของชุดข้อมูล MIRflickr

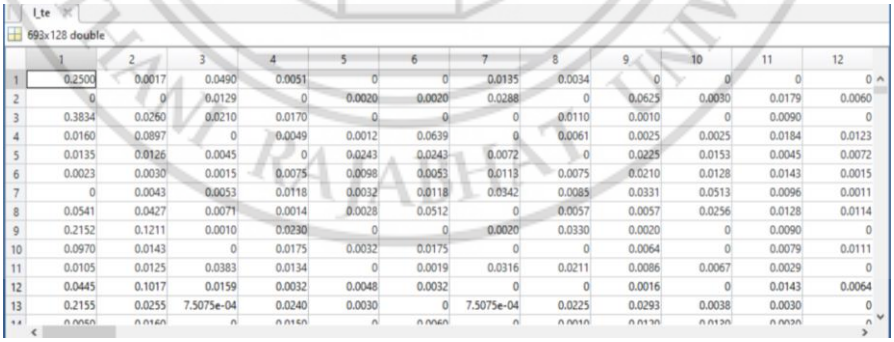


ภาพผนวกที่ 6 ตัวอย่างรายละเอียดข้อมูลข้อความของชุดข้อมูล MIRflickr

3. ชุดข้อมูล WIKI ประกอบด้วยข้อมูล 2 ประเภทคือ ข้อความและรูปภาพ โดยข้อมูลถูกแบ่งออกเป็นข้อมูลที่ใช้ทดสอบ 693 รายการ ที่เหลือเป็น reference set ข้อมูลรูปภาพมีจำนวน 128 มิติ ข้อความมีจำนวน 10 มิติ ข้อมูลลาเบลจำนวน 10 มิติ รายละเอียดต่างๆ แสดงดังภาพผนวกที่ 7-9



ภาพผนวกที่ 7 จำนวนข้อมูลของชุดข้อมูล Wiki



ภาพผนวกที่ 8 ตัวอย่างรายละเอียดข้อมูลรูปภาพของชุดข้อมูล WIKI

I_tr

2173x128 double

	1	2	3	4	5	6	7	8	9	10
1	0.0373	0.0039	0.0528	0	0.0026	0	0.0373	0	0.0090	0.0026
2	0.0204	0.0296	0.0500	0.0111	0	0	0.0333	0.0167	0.0056	0.0241
3	0	0.0035	0.0255	0.0081	0.0023	0	0.0440	0.0104	0.0069	0.0289
4	0.0139	0.0262	0.0262	0.0123	0	0	0.0123	0.0401	0	0.0082
5	0.0204	0.0188	0.0466	0.0082	0	0	0.0237	0.0090	0.0294	0.0098
6	0.0430	0.0209	0	0.0221	0.0012	0.0012	0	0	0.0283	0.0074
7	0.0373	0.0317	0.0233	0.0140	0.0028	0.0121	0.0028	0	0.0065	0.0047
8	0.0530	0.0290	0.0417	0.0025	0.0227	0.0038	0.0088	0	0.0051	0.0076
9	0.0169	0.0315	0.0023	0.0450	0	0.0158	0.0315	0.0383	0.0068	0.0101
10	0.0119	0	0.0389	0	0.0043	0.0130	0	0	0.0227	0.0076
11	0.0260	0.0120	0.0601	0.0050	0.0050	0.0320	0	0	0.0180	0
12	0	0.0231	0.0504	0	0	0	0.0231	0.0336	0.0084	0.0168
13	0.1289	0.0603	0.0187	0.0052	0	0.0353	0	0.0021	0.0260	0.0249
14	0	0	0	0	0	0.0299	0	0	0	0.0024

ภาพผนวกที่ 9 ตัวอย่างรายละเอียดข้อมูลข้อความของชุดข้อมูล WIKI

4. ชุดข้อมูล XMedia ประกอบด้วยข้อมูล 5 ประเภทคือ ข้อความ รูปภาพ วิดีโอ เสียง และข้อมูล 3 มิติ โดยข้อมูลถูกแบ่งออกเป็นข้อมูลที่ใช้ทดสอบดังนี้ 1) ข้อความ 1000 รายการ 10 มิติ 2) รูปภาพ 1000 รายการ 128 มิติ 3) วิดีโอ 174 รายการ 128 มิติ 4) เสียง 200 รายการ 29 มิติ และ 5) ข้อมูล 3 มิติ 100 รายการ 4700 มิติ ตามลำดับ โดยข้อมูลที่เหลือเป็น reference set ทุกประเภทข้อมูลมีจำนวนลาเบลที่เท่ากันคือจำนวน 10 มิติ รายละเอียดต่าง ๆ แสดงดังภาพผนวกที่ 9-14

Name	Value	Min	Max
A_te	200x29 double	-7.7260	12.8001
A_tr	800x29 double	-9.3822	14.0087
d3_te	100x4700 double	0	255
d3_tr	400x4700 double	<Too ...	<Too ...
I_te	1000x128 double	0	0.9594
I_te_CNN	1000x4096 double	<Too ...	<Too ...
I_tr	4000x128 double	<Too ...	<Too ...
I_tr_CNN	4000x4096 double	<Too ...	<Too ...
T_te	1000x10 double	0.0609	0.2463
T_te_BOW	1000x3000 double	<Too ...	<Too ...
T_tr	4000x10 double	0.0628	0.2174
T_tr_BOW	4000x3000 double	<Too ...	<Too ...
te3dCat	100x1 double	1	20
teAudCat	200x1 double	1	20
teImgCat	1000x1 double	1	20
teTxtCat	1000x1 double	1	20
teVidCat	174x1 double	1	20
tr3dCat	400x1 double	1	20
trAudCat	800x1 double	1	20
trImgCat	4000x1 double	1	20
trTxtCat	4000x1 double	1	20
trVidCat	969x1 double	1	20
V_te	174x128 single	0	1
V_te_CNN	174x4096 double	<Too ...	<Too ...
V_tr	969x128 single	0	1
V_tr_CNN	969x4096 double	<Too ...	<Too ...

ภาพผนวกที่ 9 จำนวนข้อมูลแต่ละประเภทของชุดข้อมูล XMedia

A_te

200x29 double

	1	2	3	4	5	6	7	8	9	10	11	12
1	8.4292	0.8995	-0.8911	1.6919	-0.4609	2.0239	0.3537	0.2344	1.0861	0.0845	1.2508	0.2154
2	7.4704	5.3571	-1.2295	2.8149	-0.9823	2.1680	-0.7713	1.5198	-0.3909	0.9111	0.0069	0.5293
3	8.8108	1.3783	0.5142	2.6586	1.0105	0.4502	2.6572	1.3611	-0.1822	1.4375	0.7337	-0.3118
4	10.1106	4.4998	-2.7500	3.0230	-0.9790	2.2891	-0.4006	1.1911	0.2030	0.4813	0.5345	0.0631
5	11.2228	0.9302	-4.3883	4.6174	-0.0572	0.9118	0.6557	1.5673	0.7319	-0.2087	2.0738	-0.5675
6	8.8429	5.9135	-0.7422	2.6439	-1.5842	1.6849	-0.8842	1.5110	-0.3132	0.9466	0.1902	0.5034
7	8.3782	2.2809	-1.3508	4.0131	-1.6255	2.1743	-0.2083	0.5600	1.1979	0.2693	0.9421	0.3485
8	9.5154	1.9045	-2.1187	4.2915	-0.4367	2.1696	0.4827	0.7590	0.8212	0.5012	1.2814	-0.0648
9	11.9726	-2.9013	-2.2751	3.1424	1.1311	-0.6027	2.0358	1.6441	-1.0727	0.0199	1.8006	0.6088
10	10.0700	2.1261	-4.3836	3.9857	-1.2251	1.2289	-0.1308	1.3366	0.4871	-0.0672	1.1369	-0.2432
11	7.7469	2.7998	-1.2018	-0.1816	-1.4599	0.2328	-0.4218	-0.5218	-0.5016	-1.8688	-0.8059	-1.1565
12	7.2743	-0.7663	-1.5640	2.6339	-0.4290	0.6172	0.8518	0.2142	0.8727	-0.2463	1.2108	-0.2010
13	9.0669	5.3463	-0.8431	3.4348	0.9320	2.3211	-0.0990	1.6248	0.2531	1.2915	0.8124	0.9667

ภาพผนวกที่ 10 ตัวอย่างรายละเอียดข้อมูลเสียงของชุดข้อมูล XMedia

d3_te

100x4700 double

	1	2	3	4	5	6	7	8	9	10	11	12
1	23	118	40	11	30	52	23	21	8	6	3	26
2	53	79	84	13	43	36	9	12	61	15	29	52
3	31	123	46	10	22	16	8	12	16	11	14	45
4	65	91	43	14	35	48	21	25	7	6	4	75
5	140	1	27	18	6	78	17	59	23	21	10	87
6	102	84	43	2	38	49	2	41	50	5	42	7
7	0	91	76	15	42	25	6	17	72	23	26	14
8	19	184	208	10	188	244	10	190	201	13	185	233
9	47	23	17	13	8	227	21	46	27	16	3	205
10	14	104	19	12	8	37	24	5	6	4	1	25
11	6	68	143	13	57	222	17	80	129	13	51	175
12	81	97	62	27	11	78	18	44	55	9	37	35
13	91	107	110	36	57	3	14	3	50	12	42	40

ภาพผนวกที่ 11 ตัวอย่างรายละเอียดข้อมูล 3 มิติของชุดข้อมูล XMedia

I_te

1000x128 double

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.0019	0.0032	0.4636	6.1139e-04	7.6423e-05	5.3496e-04	0.0041	0.0028	0.0091	0.0014	0.0167	3.8212e-0
2	0.0020	0.0015	0.4543	0.0067	0.0025	0.0029	0.0019	0.0055	0.0015	0.0027	0.0019	0.002
3	0.0047	0.0079	0.4054	0.0053	0.0019	0.0056	0.0022	0.0031	0.0153	0.0036	0.0047	0.003
4	0.0062	0.0021	0.4316	0.0062	0.0023	0.0023	0.0012	0.0032	0.0125	0.0082	1.5782e-04	0.005
5	0	0.0027	0.4906	3.5829e-04	3.5829e-04	5.9716e-04	0.0067	3.5829e-04	0.0061	0.0030	0.0176	5.9716e-0
6	0.0281	0.0051	0.0126	0.0087	0.0051	0.0082	0.0033	0.0047	0.0056	0.0064	0.0016	0.023
7	0.0032	0.0024	0.3652	0.0026	0.0021	0.0055	0.0028	0.0058	0.0093	0.0083	0.0026	0.001
8	0.0180	0.0052	0.1264	0.0032	0.0019	0.0053	0.0017	7.7785e-04	0.0094	0.0071	0.0033	0.015
9	0.0034	0.0026	0.2219	0.0041	0.0033	0.0041	0.0015	0.0036	0.0151	0.0051	0.0028	0.007
10	0.0106	0.0077	0.1000	0.0015	0.0020	0.0026	0.0043	7.8339e-05	0.0068	0.0092	0.0080	0.004
11	0.0072	0.0058	3.5395e-04	0.0021	0.0018	0.0055	0.0024	0.0029	0.0076	0.0075	0.0071	0.005
12	0.0012	0.0026	0.3555	9.1075e-05	0.0024	0.0015	9.1075e-05	0.0015	0.0012	0.0057	6.3752e-04	0.001
13	0.0117	0.0098	0.0044	0	0.0015	0.0026	0.0018	0	0.0016	0.0158	0	0.003

ภาพผนวกที่ 12 ตัวอย่างรายละเอียดข้อมูลรูปภาพของชุดข้อมูล XMedia

T_te

1000x10 double

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.0862	0.0783	0.0994	0.1099	0.0697	0.1122	0.0816	0.0981	0.0976	0.1669		
2	0.0997	0.0882	0.1127	0.1142	0.0788	0.1118	0.0954	0.0877	0.1011	0.1105		
3	0.1006	0.0823	0.1132	0.1194	0.0775	0.1134	0.0908	0.0827	0.1094	0.1107		
4	0.0962	0.0884	0.1139	0.1114	0.0752	0.1078	0.1096	0.0896	0.1061	0.1018		
5	0.0981	0.0856	0.1154	0.1154	0.0774	0.1122	0.0921	0.0871	0.1071	0.1096		
6	0.0990	0.0846	0.1122	0.1192	0.0726	0.1135	0.0921	0.0858	0.1095	0.1115		
7	0.1005	0.0826	0.1173	0.1223	0.0734	0.1112	0.0927	0.0845	0.1028	0.1128		
8	0.0988	0.0875	0.1170	0.1156	0.0735	0.1119	0.0935	0.0855	0.1121	0.1046		
9	0.1012	0.0854	0.1118	0.1192	0.0755	0.1064	0.0922	0.0949	0.1022	0.1112		
10	0.0944	0.0828	0.1207	0.1183	0.0735	0.1140	0.0997	0.0841	0.1070	0.1054		
11	0.0979	0.0860	0.1128	0.1185	0.0794	0.1124	0.0929	0.0864	0.1046	0.1092		
12	0.1014	0.0882	0.1118	0.1150	0.0753	0.1145	0.0927	0.0867	0.1043	0.1102		
13	0.1001	0.0904	0.1089	0.1170	0.0718	0.1114	0.0907	0.0865	0.1101	0.1131		

ภาพผนวกที่ 13 ตัวอย่างรายละเอียดข้อมูลข้อความของชุดข้อมูล XMedia

T_tr_BOW

4000x3000 double

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0.0238	0	0	0	0	0	0.023
4	0	0	0	0	0	0	0	0	0	0	0	
5	0	0	0	0.0769	0	0	0	0.0769	0	0	0	
6	0	0	0	0	0	0	0	0	0	0	0	
7	0.0196	0	0	0	0	0	0	0	0.0196	0	0	0.019
8	0.0769	0	0	0	0	0	0	0	0	0	0	
9	0	0	0	0	0	0	0	0	0	0	0	
10	0	0	0	0	0	0	0	0	0.0128	0	0	
11	0	0	0	0	0	0	0	0	0	0	0	
12	0	0	0	0	0	0	0	0	0	0	0	
13	0	0	0	0	0	0	0	0	0	0	0	

ภาพผนวกที่ 14 ตัวอย่างรายละเอียดข้อมูลวิดีโอของชุดข้อมูล XMedia

ภาคผนวก ข ประวัติผู้วิจัย

ชื่อ-นามสกุล:

ดร.ศราวุธ มากชิต (Sarawut Markchit, Ph.D)

ติดต่อ:

อีเมล: sarawut.mar@sru.ac.th

หมายเลขโทรศัพท์: 08-6694-3135

ประวัติการศึกษา:

มัธยมศึกษา: โรงเรียนพัทลุงพิทยาคม

ประกาศนียบัตร: ประกาศนียบัตรวิชาชีพครู มหาวิทยาลัยสุโขทัยธรรมาธิราช

ปริญญาตรี: วท.บ. วิทยาการคอมพิวเตอร์ มหาวิทยาลัยสงขลานครินทร์

ปริญญาโท: วท.ม. การจัดการเทคโนโลยีสารสนเทศ มหาวิทยาลัยวลัยลักษณ์

ปริญญาเอก: ปร.ด. วิทยาการคอมพิวเตอร์และวิศวกรรมสารสนเทศ (Ph.D. Computer Science and Information Engineering), National Chiayi University, ไต้หวัน