บทที่ 4

การทดลองและผลการทดลอง

การทดลองในบทนี้มีวัตถุประสงค์เพื่อทำการทดสอบประสิทธิภาพด้วยการวัดความเร็วของขั้นตอนวิธีระหว่างขั้นตอนวิธี MCCRT (Minimum Classification Correction Rate Transformation Algorithm) กับขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้น (Incremental Algorithm) เนื่องจากในบทที่ 3 ได้แสดงให้เห็นแล้วว่าค่าความซับซ้อนเชิงคำนวณของขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นนั้นน้อยกว่าขั้นตอนวิธีแบบ MCCRT ในหลายกรณี ซึ่งในการทดลองต้องการทราบว่าเมื่อตัวแปรต่างๆของปัญหามีการเปลี่ยนแปลง แนวโน้มของประสิทธิภาพจะเป็นไปดังความซับซ้อนเชิงคำนวณในบทที่ 3 หรือไม่ โดยตัวแปรที่จะพิจารณามีดังต่อไปนี้ ตัวแปรแรกคือ $|\Delta d|$ หรือขนาดของข้อมูลที่เพิ่มขึ้นกล่าวคือจำนวนระเบียนที่เพิ่มเข้ามาของข้อมูล ตัวแปรต่อมาคือค่า k ใน k-Anonymity และตัวแปรสุดท้ายคือ $|Q_D|$ หรือจำนวนของคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อม

4.1 การเตรียมการทดลอง

ข้อมูลที่ใช้ในการทดลองคือ ข้อมูล CRX ซึ่งมีระเบียนเป็นจำนวน 6710 ระเบียนและข้อมูล Breast-Cancer ซึ่งมีระเบียนเป็นจำนวน 699 ระเบียนข้อมูลทั้งสองข้อมูลมาจาก UCI Repository [6] ทั้งสองข้อมูลจะถูกทำการประมวลผลล่วงหน้า (Pre-Processing) โดยการคัดเลือกคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อม การแทนค่าข้อมูลที่ข้อมูลขาดหายไป และแบ่งแยกข้อมูล (Discrete) ให้กับข้อมูลในคอลัมน์ที่มีค่าข้อมูลเป็นค่าต่อเนื่อง (Continuous) ข้อมูลที่ได้จากการประมวลผลล่วงหน้าทั้งสองข้อมูลจะถูกแบ่งออกเป็น 2 ส่วนโดยวิธีการสุ่มระเบียน ส่วนที่หนึ่งมีขนาดเป็น 50% ของข้อมูลทั้งหมดใช้เป็นข้อมูลเริ่มต้น D ส่วนที่สองมีขนาดข้อมูลเป็น 50% ใช้เป็นข้อมูลเพิ่มเข้ามา ΔD และข้อมูลที่เพิ่มเข้ามานี้จะถูกแบ่งออกเป็นข้อมูลหลายข้อมูลโดยวิธีการสุ่มระเบียนเช่นกัน ซึ่งจำนวนการแบ่งนั้นขึ้นอยู่กับหัวข้อการวัดผลการทดลอง ตัวอย่างเช่นหัวข้อการทดลองกำหนดให้ขนาดของข้อมูลที่ถูกเพิ่มขึ้นเท่ากับ 5% ของข้อมูลเริ่มต้น D ดังนั้น ΔD จะถูกแบ่งออกเป็น 20 ข้อมูลเท่าๆกัน ซึ่งจะได้ $\Delta D1, \Delta D2, \ldots, \Delta D20$ ซึ่งในที่นี้เรียกข้อมูลที่ถูกเพิ่มขึ้นใดๆที่แบ่งแล้วนี้ว่า ΔDn

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

การทดลองในงานวิจัยนี้จะเริ่มจากการจัดเตรียมข้อมูลดังได้กล่าวมาข้างต้นเป็นจำนวน 5 ชุดข้อมูลของ CRX และ 5 ชุดข้อมูลของ Breast เพื่อกระจายความเป็นไปได้ของการเพิ่มข้อมูลที่มีค่าข้อมูลเหมือนกันหรือการเพิ่มข้อมูลที่มีค่าข้อมูลที่เป็นค่าใหม่ออกไปตาม ΔDn ที่แตกต่างกัน และการวัดเวลาจะเป็นการหาค่าเฉลี่ยของเวลาในการผ่านขั้นตอนวิธี MCCRT และขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้น ของข้อมูลทั้ง 5 ชุดข้อมูล

ทุกการทดลองนั้นจะวัดค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธีทั้งสองขั้นตอนวิธีเมื่อเพิ่มข้อมูลในแต่ละครั้งซึ่งทุกครั้งจะเพิ่มข้อมูล ΔDn เข้าไป 1 ชุดโดยในการทดลองของขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นจะเริ่มต้นจากการนำข้อมูลเริ่มต้น D ผ่านขั้นตอนวิธี MCCRT เพื่อเก็บค่าอัตราความแม่นยำในการจำแนก *CCR* ของแต่ละคอลัมน์ เก็บค่าระดับเจนเนอรัลไลเซชัน *GL* ที่ทำให้ข้อมูลมีคุณสมบัติ *k*-Anonymity ตามกระบวนการ MCCRT และเก็บข้อมูลหลังการแปลง D' ซึ่งเป็นข้อมูลที่มีคุณสมบัติ *k*-Anonymity ตามขั้นตอนวิธี MCCRT ค่าเหล่านี้จะถูกนำไปเป็นค่าระดับการเจนเนอรัลไลเซชันและข้อมูลนำเข้าในขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นร่วมกับข้อมูลที่เพิ่มขึ้น ΔDn หลังจากเตรียมค่าและข้อมูลนำเข้า จะเริ่มขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นและวัดเวลาที่ใช้ของขั้นตอนวิธีการแปลงข้อมูลแบบเพิ่มขึ้น เมื่อจบขั้นตอนวิธีดังกล่าวจะได้ค่าอัตราความแม่นยำในการทำนายค่าคลาส *CCR* (Classification Correction Rate) ของแต่ละคอลัมน์ ค่าระดับการเจนเนอรัลไลเซชัน *GL* (Generalization Level) ที่ทำให้ข้อมูลมีคุณสมบัติ *k*-Anonymity และข้อมูลหลังการแปลง D' ซึ่งเป็นข้อมูลที่มีคุณสมบัติ *k*-Anonymity แล้ว ค่าและข้อมูลที่ได้จากขั้นตอนวิธีแปลงข้อมูลแบบเพิ่มขึ้นของการทำงานในรอบนี้จะถูกนำไปรวมกับข้อมูลที่เพิ่มขึ้น ΔDn ข้อมูลถัดไป เพื่อเป็นค่าและข้อมูลนำเข้าของขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นในรอบถัดไป กระบวนการจะทำซ้ำไปจนครบ ΔDn ที่ถูกเตรียมไว้ แล้วจึงทำการรวมเวลาที่ใช้ไปทั้งหมดตั้งแต่การเพิ่มข้อมูลครั้งแรกจนครั้งสุดท้ายเป็นเวลารวม เริ่มกระบวนการทั้งหมดใหม่กับชุดข้อมูลเหลืออีก 4 ชุด จากนั้นหาค่าเฉลี่ยของเวลารวมของชุดข้อมูลทั้งหมดได้เป็นเวลาที่ใช้ในการทำงานของขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นเพื่อเปรียบเทียบกับเวลาของขั้นตอนวิธี MCCRT

ในการทดลองโดยใช้ขั้นตอนวิธีแบบ MCCRT จะเริ่มต้นด้วยการนำข้อมูลเริ่มต้น D รวมกับข้อมูลที่เพิ่มขึ้นมา ΔDn แล้วนำข้อมูลที่รวมกันของทั้งสองข้อมูลไปผ่านขั้นตอนวิธี MCCRT อีกครั้งแล้ววัดเวลาที่ใช้ในขั้นตอนวิธี MCCRT ข้อมูลที่ถูกรวมแล้วข้างต้นจะถูกนำมาเป็นข้อมูลเริ่มต้น D ใหม่เพื่อนำมารวมกับ ΔDn ข้อมูลถัดไป กระบวนการจะทำซ้ำไปจนครบ ΔDn ที่ถูกเตรียมไว้ ทำการรวมเวลาที่ใช้ทั้งหมดตั้งแต่การเพิ่มข้อมูลครั้งแรกจนครั้งสุดท้ายได้เป็นเวลารวม

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

เริ่มกระบวนการทั้งหมดใหม่กับชุดข้อมูลที่เหลืออีก 4 ชุดข้อมูล หาค่าเฉลี่ยของเวลารวมจากทั้ง 5 ชุดข้อมูลได้เป็นเวลาที่ใช้ในการทำงานของขั้นตอนวิธี MCCRT เพื่อใช้ในการเปรียบเทียบกับขั้นตอนวิธีการแปลงข้อมูลแบบเพิ่มขึ้น

งานวิจัยนี้ต้องการวัดประสิทธิภาพระหว่างขั้นตอนวิธี MCCRT กับขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้น โดยการวัดเวลาที่ใช้เมื่อมีข้อมูลเพิ่มขึ้นมา และงานวิจัยนี้ยังต้องการศึกษาผลกระทบจากตัวแปรในกระบวนการที่มีผลกระทบกับเวลาในการทำงานของขั้นตอนวิธีทั้งสองว่ามีผลกระทบกับขั้นตอนวิธีทั้งสองอย่างไร โดยตัวแปรที่จะศึกษาได้แก่ ค่า k หรือจำนวนระเบียนที่มีค่าของข้อมูลเหมือนกันตามคุณสมบัติ k-Anonymity ค่าจำนวนคอลัมน์ $|Q_D|$ ที่เป็นตัวบ่งชี้บุคคลทางอ้อม และขนาดของข้อมูลที่เพิ่มขึ้น $|\Delta D|$

และกำหนดให้แต่ละการทดลองมีค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำที่ใช้ในการหาค่าความแม่นยำในการทำนายคลาส CCR ของแต่ละคอลัมน์ให้กับขั้นตอนวิธี MCCRT และขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นเท่ากับ 0 เพื่อให้ได้ความแม่นยำแบบละเอียดมากที่สุดโดยคำนวณทุกกฎที่เป็นไปได้

งานวิจัยนี้ได้ทำการทดลองเปรียบเทียบประสิทธิภาพในการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีการแปลงค่าข้อมูลแบบขึ้นตามหัวข้อดังต่อไปนี้

4.1.1 การทดลองเปรียบเทียบประสิทธิภาพในการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นเมื่อค่า *k* ถูกเปลี่ยนแปลง

ในหัวข้อการทดลองนี้ได้กำหนดค่าขนาดของข้อมูลที่เพิ่มเข้ามา ΔDn เท่ากับข้อมูลละ 5% ของข้อมูลเริ่มต้นเป็นจำนวน 20 ชุดข้อมูล และให้จำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมเท่ากับ 9 คอลัมน์ ซึ่งเป็นจำนวนคอลัมน์ที่มากที่สุดของข้อมูล CRX และ Breast สำหรับการเปลี่ยนแปลงค่า k จะเปลี่ยนแปลงตั้งแต่ 2 ถึง 30

การทดลองเริ่มจากกำหนดให้ $k = 2$ เมื่อเตรียมค่าและข้อมูลนำเข้าของทั้งสองขั้นตอนวิธีแล้วจะทำการเพิ่มข้อมูล ΔDn เข้าไปทีละข้อมูลเป็นจำนวน 20 ข้อมูลผ่านขั้นตอนวิธี MCCRT และขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้น วัดเวลาที่ใช้รวมทั้งหมดในการเพิ่มข้อมูลจนครบ จากนั้นเปลี่ยนค่า $k = 3$ แล้วเริ่มกระบวรการใหม่อีกครั้ง ทำซ้ำจนครบทุกค่า k ที่กำหนดไว้

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

4.1.2 การทดลองเปรียบเทียบประสิทธิภาพในการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นเมื่อจำนวนคอลัมน์ที่เป็นตัวบ่งชี้ข้อมูลทางอ้อมมีการเปลี่ยนแปลง

ในหัวข้อการทดลองนี้ได้กำหนดค่าขนาดของข้อมูลที่เพิ่มเข้ามา ΔDn เท่ากับ 5% ของข้อมูลเริ่มต้นเป็นจำนวน 20 ชุดข้อมูล การกำหนดค่าจำนวนคอลัมน์ได้กำหนดไว้ตั้งแต่ 2 ถึง 9 โดยทุกค่าจำนวนคอลัมน์ จะทำการสุ่มคอลัมน์จากข้อมูลเดิมเป็นจำนวน 5 ครั้ง ทุกค่าจำนวนคอลัมน์จะทดสอบกับค่า k 3 ค่า คือ 2, 5 และ 10

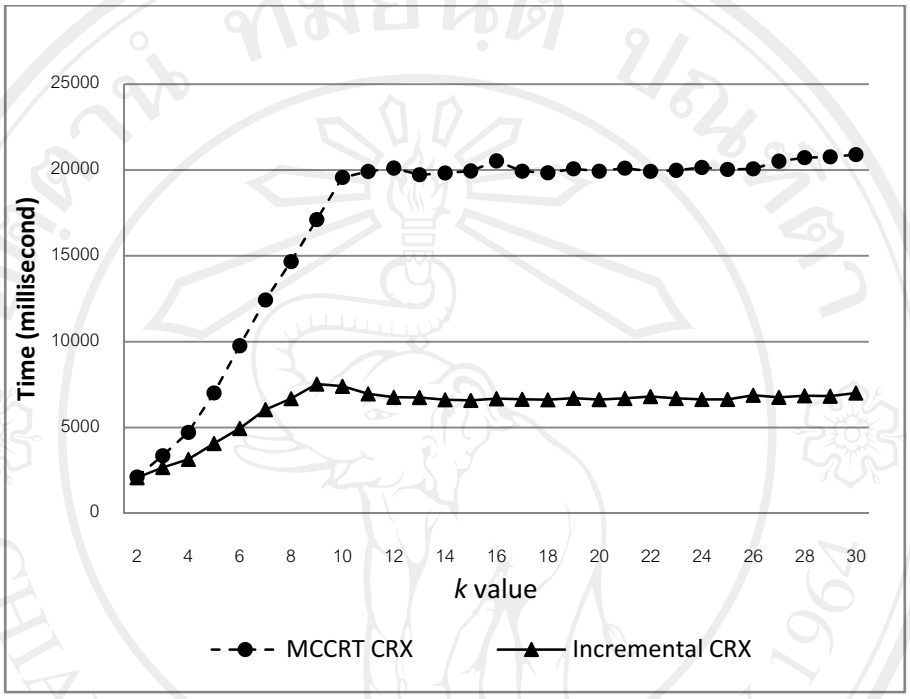
4.1.3 การทดลองเปรียบเทียบประสิทธิภาพในการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นเมื่อขนาดของข้อมูลที่เพิ่มขึ้นมีการเปลี่ยนแปลง

ขนาดของข้อมูลที่เพิ่มขึ้นในการทดลองได้กำหนดไว้ 4 ระดับคือ ให้ขนาดของข้อมูลที่เพิ่มขึ้นเป็น 1%, 2.5%, 5% และ 10% ของขนาดข้อมูลเริ่มต้น โดยขนาดของข้อมูล 1% ของขนาดข้อมูลเริ่มต้นจะมีจำนวน 100 ชุดข้อมูล ขนาดข้อมูล 2.5% ของขนาดข้อมูลเริ่มต้นจะมีจำนวน 40 ชุดข้อมูล ขนาดของข้อมูล 5% ของข้อมูลเริ่มต้นจะมีจำนวน 20 ชุดข้อมูลและขนาดของข้อมูล 10% ของข้อมูลเริ่มต้นจะมีจำนวน 10 ชุดข้อมูล ซึ่งจะนำมาทดสอบกับค่า k เท่ากับ 2, 5, และ 10 และกำหนดให้จำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมของทุกข้อมูลเท่ากับ 9

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

4.2 ผลการทดลอง

4.2.1 ผลการทดลองเปรียบเทียบประสิทธิภาพในการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นเมื่อค่า *k* ถูกเปลี่ยนแปลง



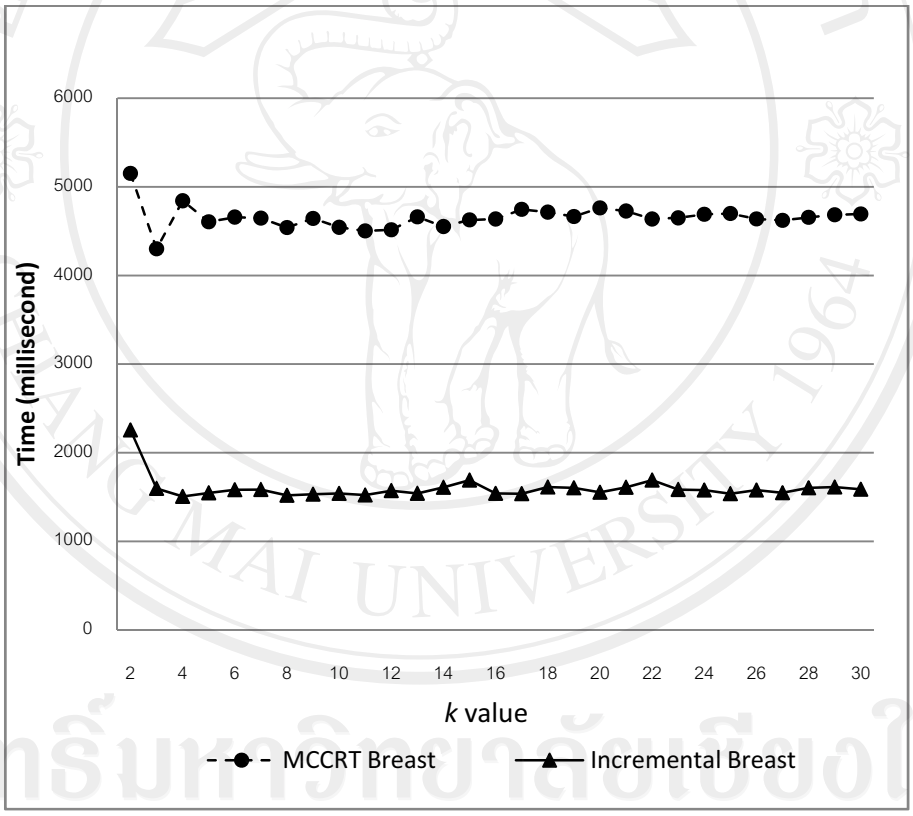
รูปที่ 4.1 ผลกระทบค่า *k* ต่อเวลาการทำงานของข้อมูล CRX

รูปที่ 4.1 ได้นำเสนอผลกระทบค่า *k* ต่อเวลาการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีแปลงค่าข้อมูลเพิ่มขึ้น ของข้อมูล CRX โดยแกนนอนคือค่า *k* ในคุณสมบัติ *k*-Anonymity และแกนตั้งคือค่าเวลาที่ใช้ในการทำงาน ค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธี MCCRT แสดงโดยเส้นกราฟ “MCCRT CRX” และค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นแสดงโดยเส้นกราฟ “Incremental CRX”

จากรูปเส้นกราฟแรกคือเส้นกราฟ “MCCRT CRX” จากจุดเริ่มต้นจะเพิ่มขึ้นเรื่อยๆจนถึง *k*=10 แล้วเส้นกราฟจะเริ่มคงที่ เส้นกราฟต่อมาคือเส้นกราฟ “Incremental CRX” จะมีพฤติกรรมที่คล้ายกับ “MCCRT CRX” แต่เส้นกราฟ “Incremental CRX” มีค่าเวลาที่ใช้ในการทำงานที่ต่ำกว่าโดยคิดเป็นค่าเฉลี่ยเพียง 30% ของ “MCCRT CRX” สาเหตุที่เป็นเช่นนี้เพราะที่ค่า *k* น้อยทำให้ข้อมูลจะมีคุณสมบัติ *k*-Anonymity ที่ระดับการเจนเนอรัลไลเซชันที่ต่ำ ซึ่งขั้นตอนวิธี MCCRT มีรอบการทำงานในการทดสอบคุณสมบัติ *k*-Anonymity ขึ้นอยู่กับระดับการเจนเนอรัลไลเซชันที่ทำให้ข้อมูลนำเข้ามีคุณสมบัติ *k*-Anonymity ดังนั้นขั้นตอนวิธี MCCRT จึงมี

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ประสิทธิภาพไม่แตกต่างกับขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นมากนัก แต่เมื่อค่า *k* เพิ่มขึ้นจะทำให้ระดับการเจนเนอรัลไลเซชันที่ทำให้ข้อมูลนำเข้ามีคุณสมบัติ *k*-Anonymity สูงขึ้นตาม ดังนั้นขั้นตอนวิธี MCCRT จึงใช้เวลามากขึ้นตามไปด้วยซึ่งในขณะที่ขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นจะใช้เวลากับการทดสอบการเปลี่ยนแปลงระดับเจนเนอรัลไลเซชันเพื่อหาจุดเริ่มต้นกับทิศทางการเปลี่ยนแปลงใหม่ แต่ใช้รอบในการทำการทดสอบคุณสมบัติ *k*-Anonymity น้อยกว่าขั้นตอนวิธี MCCRT ทำให้เมื่อค่า *k* สูงขึ้นประสิทธิภาพของขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นดีกว่าประสิทธิภาพของขั้นตอนวิธี MCCRT



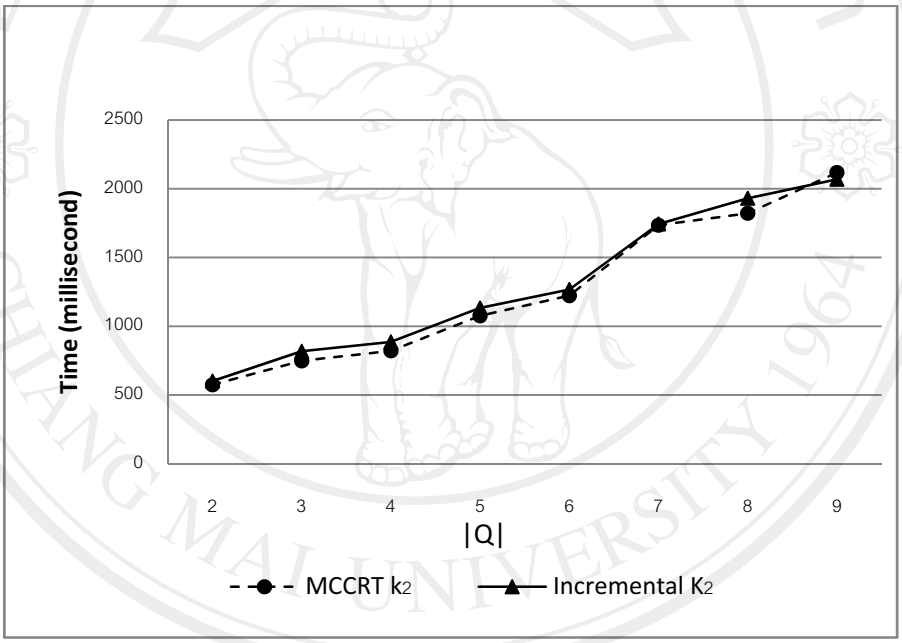
รูปที่ 4.2 ผลกระทบค่า *k* ต่อเวลาการทำงานของข้อมูล Breast

รูปที่ 4.2 ได้นำเสนอผลกระทบค่า *k* ต่อเวลาการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีแปลงค่าข้อมูลเพิ่มขึ้น ของข้อมูล Breast โดยแกนนอนคือค่า *k* ในคุณสมบัติ *k*-Anonymity และแกนตั้งคือค่าเวลาที่ใช้ในการทำงาน ค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธี MCCRT แสดงโดยเส้นกราฟ “MCCRT Breast” และค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นแสดงโดยเส้นกราฟ “Incremental Breast”

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

จากรูปเส้นกราฟ “MCCRT Breast” กับเส้นกราฟ “Incremental Breast” ค่อนข้างคงที่ตั้งแต่เริ่มต้น แต่เส้นกราฟ “Incremental Breast” มีค่าเวลาในการทำงานโดยเฉลี่ยเพียง 25% ของเส้นกราฟ “MCCRT Breast” สาเหตุเนื่องมาจากข้อมูล Breast มีระดับการเจนเนอรัลไลเซชันที่ทำให้ข้อมูลมีคุณสมบัติ k-Anonymity ในระดับสูงตั้งแต่ค่า k เท่ากับ 2 เท่านั้นทำให้ขั้นตอนวิธี MCCRT ใช้เวลาในการทำงานมากในทุกค่า k

4.2.2 ผลการทดลองเปรียบเทียบประสิทธิภาพในการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นเมื่อค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อม ($|\Delta Q|$) ในข้อมูลนำเข้าถูกเปลี่ยนแปลง



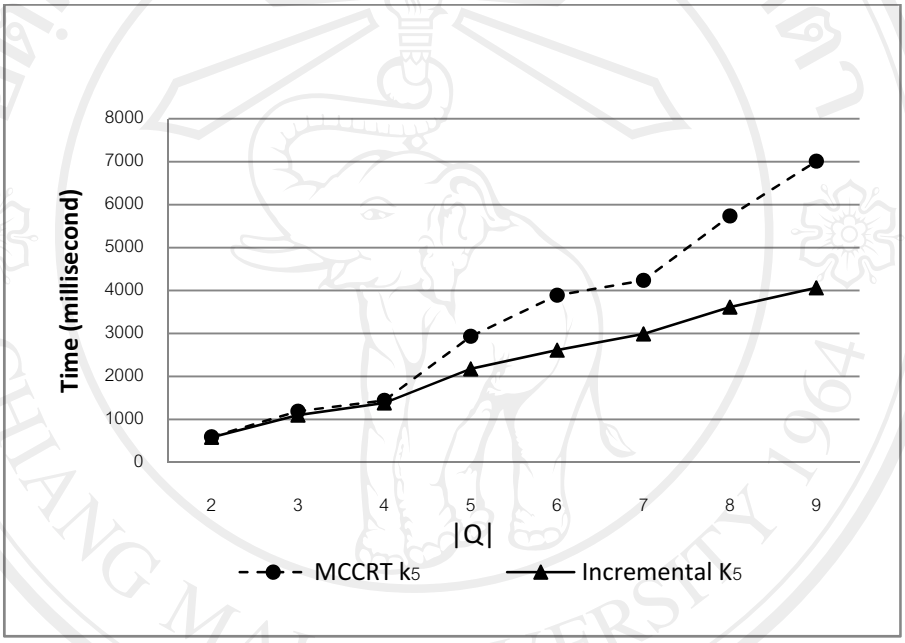
รูปที่ 4.3 ผลกระทบของค่า $|\Delta Q|$ ต่อเวลาการทำงาน ของข้อมูล CRX โดยค่า k =2

รูปที่ 4.3 ได้นำเสนอผลกระทบค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมของข้อมูลนำเข้าต่อเวลาการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีแปลงค่าข้อมูลเพิ่มขึ้น โดยขนาดข้อมูลเพิ่มขึ้นเท่ากับ 5% และมีค่า k=2 ของข้อมูล CRX โดยแกนนอนคือค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมของข้อมูลนำเข้า และแกนตั้งคือค่าเวลาที่ใช้ในการทำงาน ค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธี MCCRT แสดงโดยเส้นกราฟ “MCCRT k2” และค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นแสดงโดยเส้นกราฟ “Incremental k2”

จากรูปเส้นกราฟ “MCCRT k2” และ เส้นกราฟ “Incremental k2” จะเพิ่มขึ้นเรื่อยๆเมื่อ $|\Delta Q|$ เพิ่มขึ้นและเส้นกราฟทั้งสองมีค่าเวลาในการทำงานใกล้เคียงกันมากโดย “Incremental k2”

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

มีค่าเวลาในการทำงานมากกว่าเล็กน้อย สาเหตุเนื่องมาจาก ข้อมูล CRX มีระดับเจนเนอรัลไลเซชันที่ทำให้ข้อมูลมีคุณสมบัติ *k*-Anonymity ที่ระดับต่ำมากเมื่อ k มีค่าต่ำ ซึ่งในกรณีนี้ k =2 ทำให้ขั้นตอนวิธี MCCRT มีรอบในการทดสอบคุณสมบัติ *k*-Anonymity น้อย แต่ขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นต้องเสียเวลาในการทดสอบการลดหรือเพิ่มระดับการเจนเนอรัลไลเซชันจึงทำให้ใช้เวลาในการทำงานมากกว่าขั้นตอนวิธี MCCRT เล็กน้อย ส่วนสาเหตุการเพิ่มขึ้นของเส้นกราฟทั้งสองนั้นเนื่องจากเมื่อค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้ข้อมูลทางอ้อมเพิ่มขึ้นนั้นจะทำให้ข้อมูลที่ใช้ในการประมวลผลมีขนาดที่เพิ่มขึ้นตามไปด้วยจึงใช้เวลาในการทำงานเพิ่มขึ้นตาม



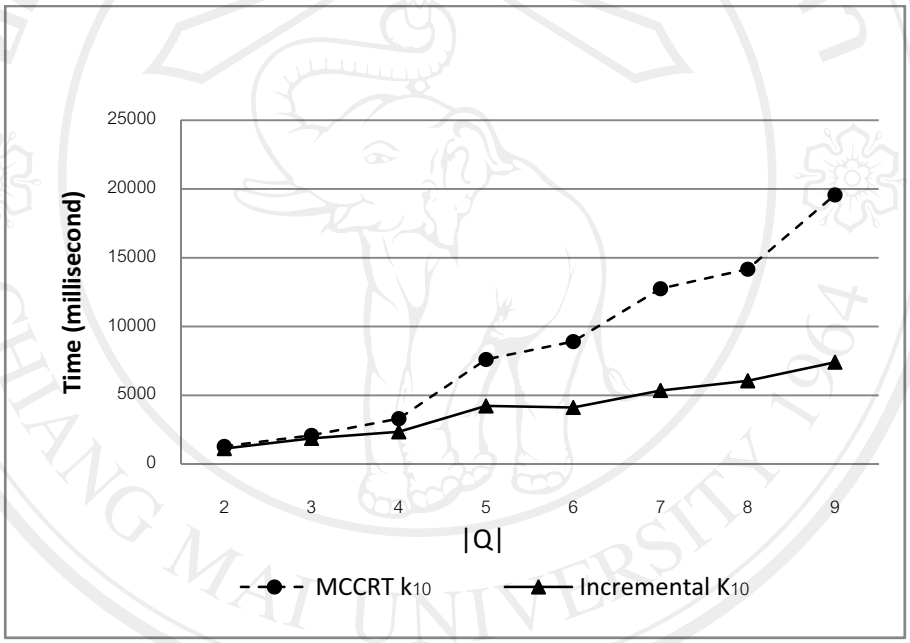
รูปที่ 4.4 ผลกระทบของค่า $|\Delta Q|$ ต่อเวลาการทำงาน ของข้อมูล CRX โดยค่า k =5

รูปที่ 4.4 ได้นำเสนอผลกระทบค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมของข้อมูลนำเข้าต่อเวลาการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีแปลงค่าข้อมูลเพิ่มขึ้น โดยขนาดข้อมูลเพิ่มขึ้นเท่ากับ 5% และมีค่า k=5 ของข้อมูล CRX โดยแกนนอนคือค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมของข้อมูลนำเข้า และแกนตั้งคือค่าเวลาที่ใช้ในการทำงาน ค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธี MCCRT แสดงโดยเส้นกราฟ “MCCRT k5” และค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นแสดงโดยเส้นกราฟ “Incremental k5”

จากรูปเส้นกราฟ “MCCRT k5” และเส้นกราฟ “Incremental k5” จะเพิ่มขึ้นเรื่อยๆเมื่อค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้ข้อมูลทางอ้อมของข้อมูลนำเข้าเพิ่มขึ้น โดยเส้นกราฟ “Incremental k5” มีค่าเวลาในการทำงานน้อยกว่าเส้นกราฟ “MCCRT k5” โดยค่าจำนวน

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

คอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมยิ่งมากมากเวลาที่ใช้ในการทำงานยิ่งแตกต่างกัน สาเหตุเนื่องมาจาก ข้อมูล CRX มีระดับเจนเนอรัลไลเซชันที่ทำให้ข้อมูลมีคุณสมบัติ *k*-Anonymity ที่ระดับสูงเมื่อ *k* มีค่าสูงซึ่งในกรณีนี้ *k* =5 ทำให้ขั้นตอนวิธี MCCRT มีรอบในการทดสอบคุณสมบัติ *k*-Anonymity มาก แต่ขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นใช้เวลาเพียงแค่ทำการทดสอบการลดหรือเพิ่มระดับการเจนเนอรัลไลเซชันจึงทำให้ใช้เวลาในการทำงานน้อยกว่าขั้นตอนวิธี MCCRT ส่วนสาเหตุที่ค่าเวลาในการทำงานเพิ่มขึ้นของเส้นกราฟทั้งสองนั้นเนื่องจากเมื่อค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้ข้อมูลทางอ้อมของข้อมูลนำเข้าเพิ่มขึ้นนั้นจะทำให้ข้อมูลที่ใช้ในการประมวลผลมีขนาดที่เพิ่มขึ้นตามไปด้วยจึงใช้เวลาในการทำงานเพิ่มขึ้นตาม



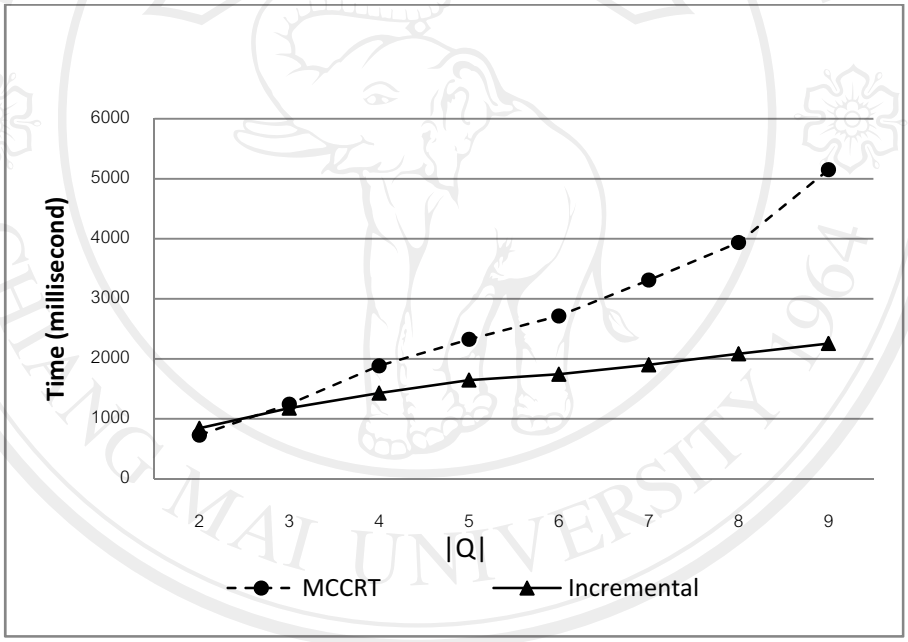
รูปที่ 4.5 ผลกระทบของค่า $|\Delta Q|$ ต่อเวลาการทำงาน ของข้อมูล CRX โดยค่า k =10

รูปที่ 4.5 ได้นำเสนอผลกระทบค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมของข้อมูลนำเข้าต่อเวลาการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีแปลงค่าข้อมูลเพิ่มขึ้น โดยมีขนาดข้อมูลเพิ่มขึ้นเท่ากับ 5% และค่า *k*=10 ของข้อมูล CRX โดยแกนนอนคือค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมของข้อมูลนำเข้า และแกนตั้งคือค่าเวลาที่ใช้ในการทำงาน ค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธี MCCRT แสดงโดยเส้นกราฟ “MCCRT k10” และค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นแสดงโดยเส้นกราฟ “Incremental k10”

จากรูปเส้นกราฟ “MCCRT k10” และเส้นกราฟ “Incremental k10” จะเพิ่มขึ้นเรื่อยๆ เมื่อค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้ข้อมูลทางอ้อมของข้อมูลนำเข้าเพิ่มขึ้น โดยเส้นกราฟ

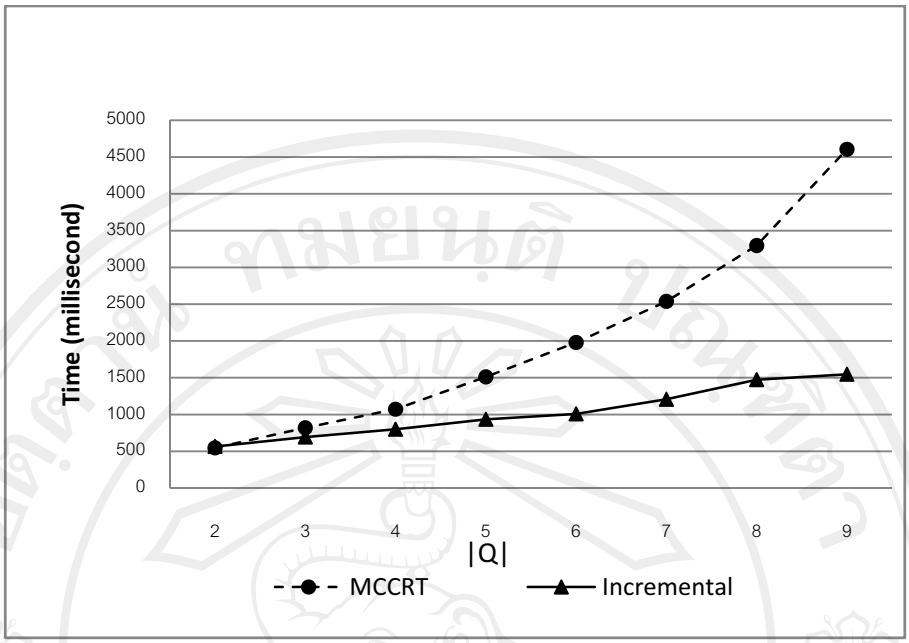
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

“Incremental k2” มีค่าเวลาในการทำงานน้อยกว่าเส้นกราฟ “MCCRT k10” มากโดยค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมยิ่งมากมากเวลาที่ใช้ในการทำงานยิ่งแตกต่างกันมาก สาเหตุเนื่องมาจาก ข้อมูล CRX มีระดับเจนเนอรัลไลเซชันที่ทำให้ข้อมูลมีคุณสมบัติ *k*-Anonymity ที่ระดับสูงเมื่อ *k* มีค่าสูงซึ่งในกรณีนี้ *k* =10 ทำให้ขั้นตอนวิธี MCCRT มีรอบในการทดสอบคุณสมบัติ *k*-Anonymity มาก แต่ขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นใช้เวลาเพียงแค่ทำการทดสอบการลดหรือเพิ่มระดับการเจนเนอรัลไลเซชันจึงทำให้ใช้เวลาในการทำงานน้อยกว่าขั้นตอนวิธี MCCRT มาก ส่วนสาเหตุที่ค่าเวลาในการทำงานเพิ่มขึ้นของเส้นกราฟทั้งสองนั้นเนื่องจากเมื่อค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้ข้อมูลทางอ้อมของข้อมูลนำเข้าเพิ่มขึ้นนั้นจะทำให้ข้อมูลที่ใช้ในการประมวลผลมีขนาดที่เพิ่มขึ้นตามไปด้วยจึงใช้เวลาในการทำงานเพิ่มขึ้นตาม

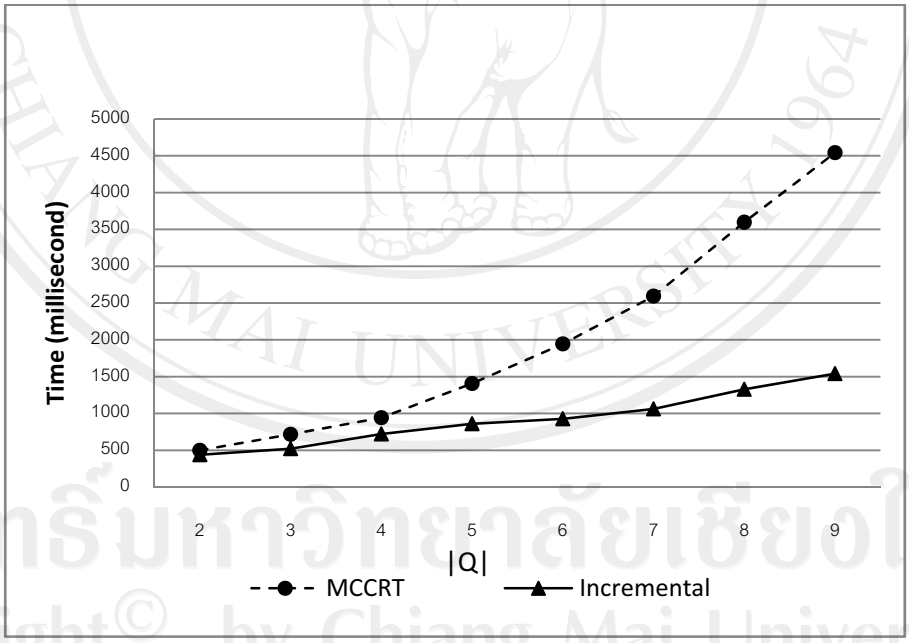


รูปที่ 4.6 ผลกระทบของค่า $|\Delta Q|$ ต่อเวลาการทำงาน ของข้อมูล Breast โดยค่า k =2

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved



รูปที่ 4.7 ผลกระทบของค่า $|\Delta Q|$ ต่อเวลาการทำงาน ของข้อมูล Breast โดยค่า k =5



รูปที่ 4.8 ผลกระทบของค่า $|\Delta Q|$ ต่อเวลาการทำงาน ของข้อมูล Breast โดยค่า k =10

รูปที่ 4.6, รูปที่ 4.7 และ รูปที่ 4.8 ได้นำเสนอผลกระทบค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมของข้อมูลนำเข้าต่อเวลาการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีแปลงค่าข้อมูลเพิ่มขึ้น โดยมีขนาดข้อมูลเพิ่มขึ้นเท่ากับ 5% และค่า k=2, k=5 และ k=10 ตามลำดับ ซึ่งมีข้อมูลนำเข้าเป็นข้อมูล Breast โดยแกนนอนคือค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมของ

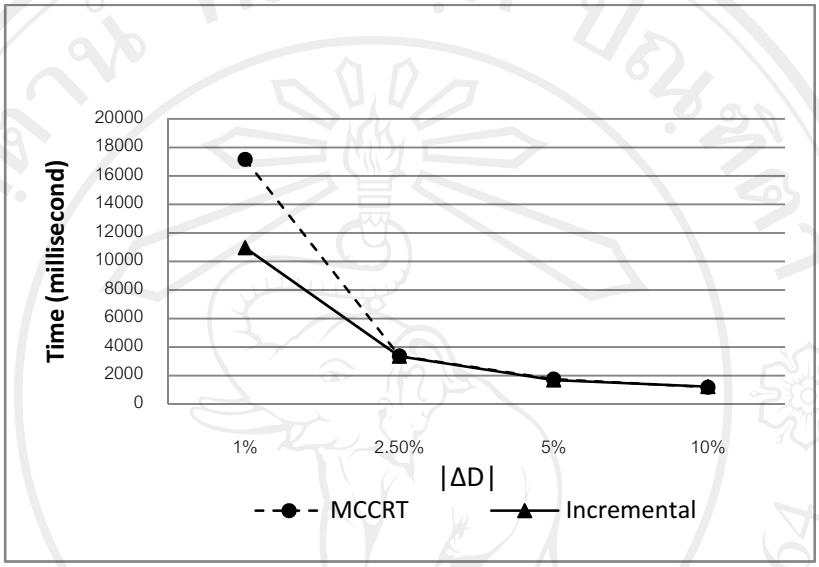
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ข้อมูลนำเข้า และแกนตั้งคือค่าเวลาที่ใช้ในการทำงาน ค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธี MCCRT แสดงโดยเส้นกราฟ “MCCRT” และค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นแสดงโดยเส้นกราฟ “Incremental”

จากสามรูปที่กล่าวมาจะเห็นว่าเส้นกราฟ “MCCRT” และเส้นกราฟ “Incremental” มีพฤติกรรมที่คล้ายกันโดยจะเพิ่มขึ้นเรื่อยๆเมื่อค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้ข้อมูลทางอ้อมของข้อมูลนำเข้าเพิ่มขึ้น โดยเส้นกราฟ “Incremental” มีค่าเวลาในการทำงานน้อยกว่าเส้นกราฟ “MCCRT” มากโดยค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมยิ่งมากมากเวลาที่ใช้ในการทำงานยิ่งแตกต่างกันมาก สาเหตุเนื่องมาจาก ข้อมูล Breast มีระดับเจนเนอรัลไลเซชันที่ทำให้ข้อมูลมีคุณสมบัติ k-Anonymity ที่ระดับสูงตั้งแต่ k มีค่าต่ำซึ่งในกรณีนี้ k ทำให้ขั้นตอนวิธี MCCRT มีรอบในการทดสอบคุณสมบัติ k-Anonymity มาก แต่ขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นใช้เวลาเพียงแค่ทำการทดสอบการลดหรือเพิ่มระดับการเจนเนอรัลไลเซชันจึงทำให้ใช้เวลาในการทำงานน้อยกว่าขั้นตอนวิธี MCCRT มาก ส่วนสาเหตุที่ค่าเวลาในการทำงานเพิ่มขึ้นของเส้นกราฟทั้งสองของทั้ง 3 รูปนั้นเนื่องจากเมื่อค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้ข้อมูลทางอ้อมของข้อมูลนำเข้าเพิ่มขึ้นนั้นจะทำให้ข้อมูลที่ใช้ในการประมวลผลมีขนาดที่เพิ่มขึ้นตามไปด้วยจึงใช้เวลาในการทำงานเพิ่มขึ้นตาม

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

4.2.3 ผลการทดลองเปรียบเทียบประสิทธิภาพในการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นเมื่อขนาดของข้อมูลที่เพิ่มขึ้น $|\Delta D|$ มีการเปลี่ยนแปลง



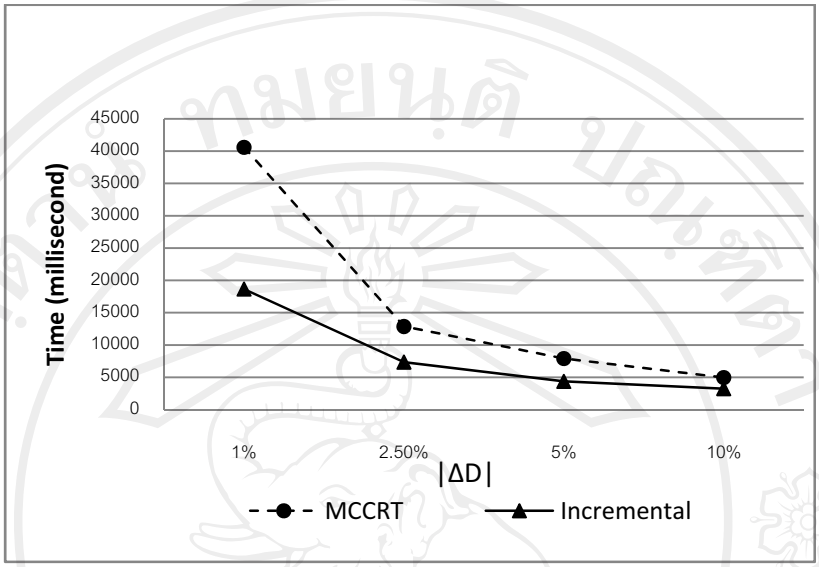
รูปที่ 4.9 ผลกระทบค่า $|\Delta D|$ ต่อเวลาในการทำงาน ของข้อมูล CRX ที่ k=2 และ $|\Delta Q|$=9

รูปที่ 4.9 ได้นำเสนอผลกระทบค่าขนาดของข้อมูลที่เพิ่มขึ้นต่อเวลาการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีแปลงค่าข้อมูลเพิ่มขึ้น โดยมีค่า k=2 และค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมเท่ากับ 9 ของข้อมูล CRX โดยแกนนอนคือค่าขนาดของข้อมูลเพิ่มขึ้น และแกนตั้งคือค่าเวลาที่ใช้ในการทำงาน ค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธี MCCRT แสดงโดยเส้นกราฟ “MCCRT” และค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นแสดงโดยเส้นกราฟ “Incremental”

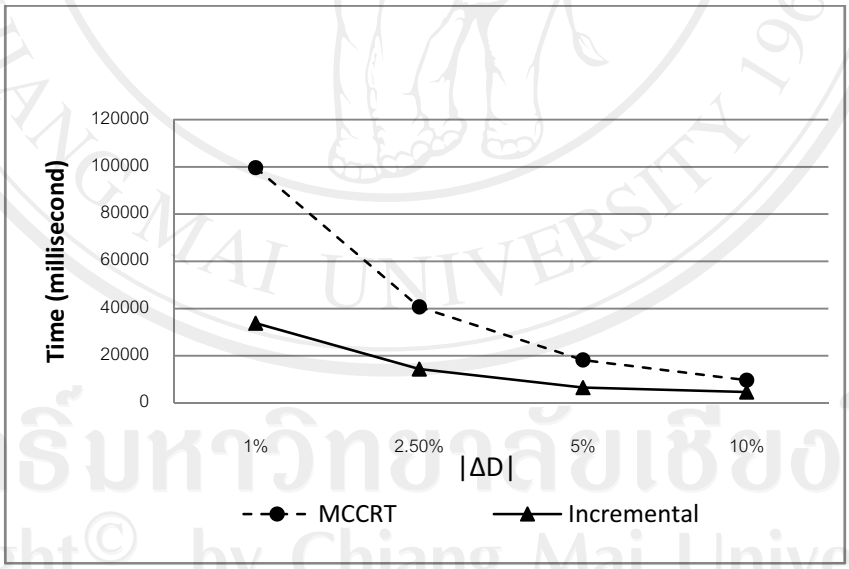
จากรูปเส้นกราฟ “MCCRT” และเส้นกราฟ “Incremental” จะลดลงเมื่อขนาดของข้อมูลเพิ่มขึ้นแต่ที่ขนาดของข้อมูลเพิ่มเข้ามีขนาดเท่ากับ 1% เส้นกราฟ “Incremental” มีค่าเวลาการทำงานที่น้อยกว่า “MCCRT” มากแต่เมื่อขนาดของข้อมูลเพิ่มขึ้นมีขนาดไม่ต่างกันเวลาที่ใช้ก็ไม่ต่างกันมากนัก สาเหตุเนื่องมาจากในการที่เพิ่มข้อมูลเพิ่มขึ้นมีขนาดที่น้อยในการทำงานรอบแรกๆของขั้นตอนวิธี MCCRT กับข้อมูลเริ่มต้นมีระดับการเจนเนอรัลไลเซชันที่ทำให้ข้อมูลมีคุณสมบัติ k-Anonymity ที่ระดับสูงเมื่อเพิ่มข้อมูลเข้าไปทีละน้อยทำให้ต้องเพิ่มข้อมูลหลายครั้งจึง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

จะทำให้ระดับการเจนเนอรัลไลเซชันที่ทำให้ข้อมูลมีคุณสมบัติ *k*-Anonymity ทำให้ขั้นตอนวิธี MCCRT ใช้เวลานาน



รูปที่ 4.10 ผลกระทบค่า $|\Delta D|$ ต่อเวลาในการทำงาน ของข้อมูล CRX ที่ k=5 และ $|\Delta Q|$=9



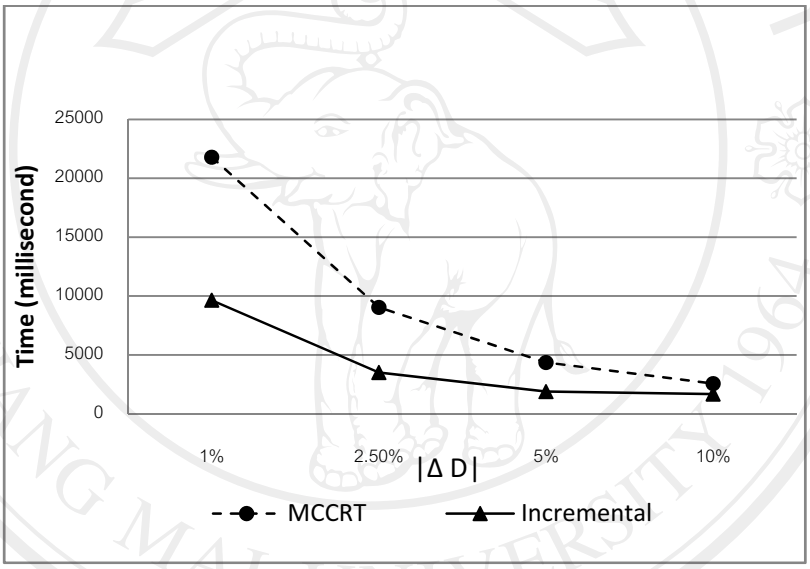
รูปที่ 4.11 ผลกระทบค่า $|\Delta D|$ ต่อเวลาในการทำงาน ของข้อมูล CRX ที่ k=10 และ $|\Delta Q|$=9

รูปที่ 4.10 และรูปที่ 4.11 ได้นำเสนอผลกระทบค่าขนาดของข้อมูลที่เพิ่มขึ้นต่อเวลาการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีแปลงค่าข้อมูลเพิ่มขึ้น โดยมีค่า k=5 และ k=10 ตามลำดับ และค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมเท่ากับ 9 ของข้อมูล CRX โดยแกนนอนคือค่าขนาดของข้อมูลเพิ่มขึ้น และแกนตั้งคือค่าเวลาที่ใช้ในการทำงาน ค่าเวลาที่ใช้ในการ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

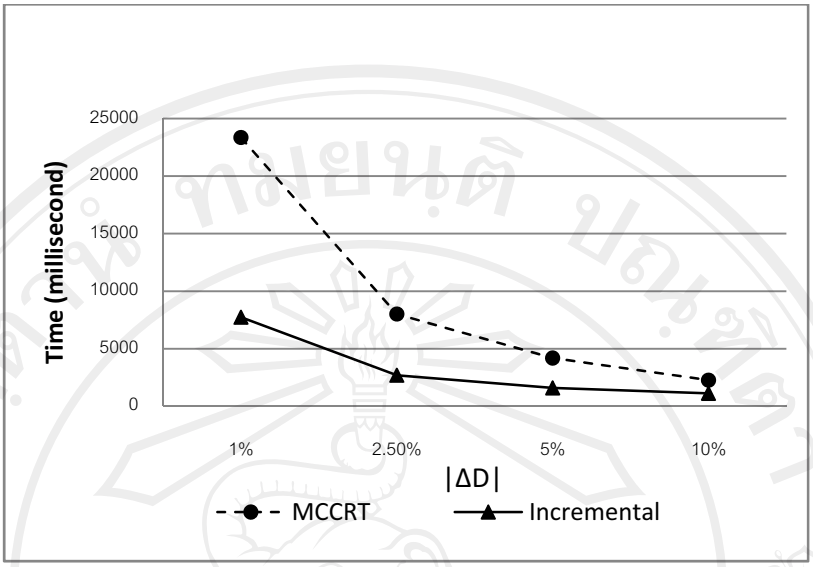
ทำงานของขั้นตอนวิธี MCCRT แสดงโดยเส้นกราฟ “MCCRT” และค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นแสดงโดยเส้นกราฟ “Incremental”

จากรูปทั้งสองจะเห็นว่าจะมีพฤติกรรมคล้ายกับเมื่อ k มีขนาดที่เพิ่มขึ้น โดยเส้นกราฟ “MCCRT” และเส้นกราฟ “Incremental” จะใช้เวลาในการทำงานลดลงเมื่อขนาดของข้อมูลเพิ่มขึ้นมีค่ามากขึ้น แต่เส้นกราฟ MCCRT จะมีค่าเวลาในการทำงานที่มากกว่าของเส้นกราฟ “Incremental” และระยะห่างเวลาของทั้งสองกราฟจะค่อยๆลดลงเมื่อขนาดของข้อมูลเพิ่มขึ้นมีขนาดมากขึ้น สาเหตุเนื่องมาจากเมื่อขนาดของข้อมูลเพิ่มขึ้นมีขนาดใหญ่มากขึ้นในการทดลองทำให้จำนวนรอบในการเพิ่มข้อมูลลดลงทำให้ขั้นตอนวิธีทั้งสองมีจำนวนรอบการทำงานลดลงไปด้วย

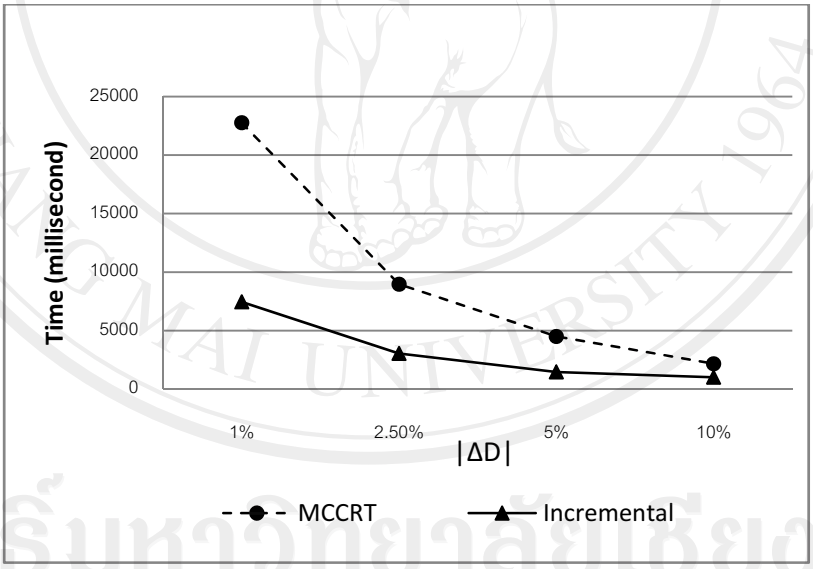


รูปที่ 4.12 ผลกระทบค่า $|\Delta D|$ ต่อเวลาในการทำงาน ของข้อมูล Breast ที่ k=2 และ $|\Delta Q|$=9

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved



รูปที่ 4.13 ผลกระทบค่า $|\Delta D|$ ต่อเวลาในการทำงาน ของข้อมูล Breast ที่ k=5 และ $|\Delta Q|$=9



รูปที่ 4.14 ผลกระทบค่า $|\Delta D|$ ต่อเวลาในการทำงาน ของข้อมูล Breast ที่ k=10 และ $|\Delta Q|$=9

รูปที่ 4.12 รูปที่ 4.13 และรูปที่ 4.14 ได้นำเสนอผลกระทบค่าขนาดของข้อมูลที่เพิ่มขึ้นต่อเวลาการทำงานของขั้นตอนวิธี MCCRT กับขั้นตอนวิธีแปลงค่าข้อมูลเพิ่มขึ้น โดยมีค่า k=2, k=5 และ k=10 ตามลำดับ และค่าจำนวนคอลัมน์ที่เป็นตัวบ่งชี้บุคคลทางอ้อมเท่ากับ 9 ของข้อมูล Breast โดยแกนนอนคือค่าขนาดของข้อมูลเพิ่มขึ้น และแกนตั้งคือค่าเวลาที่ใช้ในการทำงาน ค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธี MCCRT แสดงโดยเส้นกราฟ “MCCRT” และค่าเวลาที่ใช้ในการทำงานของขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นแสดงโดยเส้นกราฟ “Incremental”

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

จากรูปที่ 4.12 รูปที่ 4.13 และรูปที่ 4.14 จะเห็นว่าจะมีพฤติกรรมคล้ายกับแม้ k มีขนาดที่ต่างกัน โดยเส้นกราฟ “MCCRT” และเส้นกราฟ “Incremental” จะใช้เวลาในการทำงานลดลงเมื่อขนาดของข้อมูลเพิ่มขึ้นมีค่ามากขึ้น แต่เส้นกราฟ MCCRT จะมีค่าเวลาในการทำงานที่มากกว่าของเส้นกราฟ “Incremental” และระยะห่างเวลาของทั้งสองกราฟจะค่อยๆลดลงเมื่อขนาดของข้อมูลเพิ่มขึ้นมีขนาดมากขึ้น สาเหตุเนื่องมาจากเมื่อขนาดของข้อมูลเพิ่มขึ้นมีขนาดใหญ่มากขึ้นในการทดลองทำให้จำนวนรอบในการเพิ่มข้อมูลลดลงทำให้ขั้นตอนวิธีทั้งสองมีจำนวนรอบการทำงานลดลงไปด้วย และอีกสาเหตุเนื่องมาจากข้อมูล Breast มีระดับเจนเนอรัลไลเซชันที่ทำให้ข้อมูลมีคุณสมบัติ k-Anonymity สูงกับทุกค่า k

4.3 สรุปผลการทดลอง

จากผลการทดลองทุกหัวข้อการทดลองจะเห็นว่ากรณีที่ขั้นตอนวิธี MCCRT มีประสิทธิภาพในการทำงานใกล้เคียงกับขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นคือ กรณีที่ข้อมูลมีระดับการเจนเนอรัลไลเซชันที่ทำให้ข้อมูลมีคุณสมบัติ k-Anonymity ที่ต่ำมากกับค่า k ที่น้อย ซึ่งระยะห่างของเวลาที่ใช้ในการทำงานไม่ห่างกันมากนักคิดโดยเฉลี่ยเป็น 5% และในกรณีดังกล่าวเมื่อขนาดของข้อมูลเพิ่มขึ้นมีค่าน้อยมากๆ กลับทำให้ประสิทธิภาพในการทำงานของขั้นตอนวิธี MCCRT น้อยกว่าขั้นตอนวิธีแปลงค่าข้อมูลแบบเพิ่มขึ้นและในกรณีอื่นๆนั้นจะเห็นได้ว่าขั้นตอนวิธีการแปลงค่าข้อมูลแบบเพิ่มขึ้นมีประสิทธิภาพสูงกว่าขั้นตอนวิธี MCCRT 3 เท่าโดยประมาณ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved