

# การเตรียมข้อมูลเพื่อการทำเหมืองกระบวนการจากอุปกรณ์เซ็นเซอร์

## Data Preparation for Process Mining based on Sensing Devices

เอนก นามขันธ์<sup>1</sup> ภูริเดช อากาศัตย์<sup>2</sup> ประจัน พลังสันติกุล<sup>3</sup> นรณัฐ สงวนศักดิ์โยธิน<sup>4</sup>

สมพงษ์ ตุ่มสวัสดิ์<sup>5</sup> วิเชียร เปรมชัยสวัสดิ์<sup>6</sup> และนุชรี เปรมชัยสวัสดิ์<sup>7</sup>

บัณฑิตวิทยาลัย สาขาเทคโนโลยีสารสนเทศ มหาวิทยาลัยสยาม<sup>1,2,3,4,5,6</sup>

วิทยาลัยครีเอทีฟดีไซน์แอนด์เอ็นเตอร์เทนเมนต์ มหาวิทยาลัยธุรกิจบัณฑิต<sup>7</sup>

E-mail: anm\_pum@hotmail.com<sup>1</sup> poo\_arp@siam.edu<sup>2</sup> prajin.ast@gmail.com<sup>3</sup>

### บทคัดย่อ

งานวิจัยนี้นำเสนอเทคนิคในการเตรียมข้อมูลที่ถูกบันทึกจากอุปกรณ์เซ็นเซอร์เพื่อใช้ในการทำเหมืองกระบวนการ เนื่องจากในปัจจุบันมีอุปกรณ์เซ็นเซอร์อยู่มากมายไม่ว่าจะอยู่ในรูปของอินเทอร์เน็ตของสรรพสิ่ง การระบุตำแหน่งโดยใช้ Wi-Fi โดยเฉพาะอย่างยิ่งจากเซ็นเซอร์ที่มีอยู่ในโทรศัพท์มือถือ ซึ่งเป็นอุปกรณ์พกพาที่คนส่วนใหญ่มักพกติดตัวอยู่ตลอดเวลาการได้มาซึ่งข้อมูลจากเซ็นเซอร์เหล่านี้ย่อมทำให้ได้ข้อมูลจริงที่เกี่ยวข้องกับพฤติกรรมของบุคคลนั้นๆ เนื่องจากความรู้และความเข้าใจในพฤติกรรมมนุษย์เป็นสิ่งที่มีความสำคัญอย่างมากในการนำไปประยุกต์ใช้ ทั้งในเรื่องของสุขภาพ การแพทย์ การศึกษา และอื่นๆ อีกมากมายกิจกรรมของแต่ละบุคคลย่อมมีผลกระทบต่อตนเอง ผู้ที่อยู่รอบตัวเขาหรือเธอ สังคมและสิ่งแวดล้อมการวิจัยนี้จึงเน้นไปที่เทคนิคในการเตรียมข้อมูลเพื่อการทำเหมืองกระบวนการกับข้อมูลที่รวบรวมมาได้จากเซ็นเซอร์เนื่องจากเป็นขั้นตอนที่มีความสำคัญอย่างมากในการวิเคราะห์ข้อมูลบทความนี้ได้แสดงขั้นตอนในการเตรียมบันทึกเหตุการณ์จากอุปกรณ์เซ็นเซอร์ พร้อมตัวอย่างการศึกษาจากกิจกรรม การเดิน การวิ่ง การ

อยู่นิ่ง ของแต่ละบุคคล เพื่อแสดงให้เห็นพฤติกรรมของบุคคล นอกจากนั้นยังสามารถขยายการวิเคราะห์ออกไปให้กว้างขวางมากขึ้นโดยการวิเคราะห์พฤติกรรมของบุคคลโดยการรวมเอาเซ็นเซอร์อื่นๆ เข้าไปด้วยได้ เช่น การระบุตำแหน่งที่อยู่จาก Wi-Fi เป็นต้น

### Abstract

The paper presents a data preparation technique for process mining based on the sensing devices data previously received and collected through an information system. At the present, there are many sensor devices in many forms such as Internet of Things (IoT), Wi-Fi and mobile phones. One of the advantages of smart phones is that nowadays there are multiple sensors in mobile phones, which are installed and embedded in the devices, so that people can carry them with themselves to everywhere. Another advantage of such data is that the data gathered from mobile devices can show real behavior of individual person in a very cheap

and reasonable manner. Gaining knowledge to understand human behavior can be very useful in many application areas such as healthcare, education and so on. Obviously, the activity of every individual person effects oneself, other people around them, society and environment. This research is focused on preparing data previously collected from smart mobile phones for process mining and process discovery purposes. In this study, the main objective of the works was to demonstrate the usefulness of the process mining techniques to investigate and analyze the individual activities that are limited only to walking, stationary and running. The analysis can be extended to analysis of behavior of each individual by combination of other sensors such as Wi-Fi to include the location of each individual, which is going to be presented in future works.

## 1. บทนำ

ความรู้และความเข้าใจในพฤติกรรมมนุษย์เป็นสิ่งที่มีประโยชน์อย่างมากในการนำไปประยุกต์ใช้ทั้งในเรื่องของสุขภาพ การแพทย์ การศึกษา และอื่นๆ อีกมากมาย เนื่องจากกิจกรรมของแต่ละบุคคลย่อมมีผลกระทบต่อตนเอง ผู้ที่อยู่รอบตัวเขาหรือเธอ สังคมและสิ่งแวดล้อม ในอดีตทฤษฎีเกี่ยวกับพฤติกรรมของมนุษย์และกิจกรรมต่างๆ มีรากฐานจากการเฝ้าสังเกตซึ่งเป็นไปได้อย่างจำกัดแต่ในปัจจุบันได้มีการพัฒนาการของเทคโนโลยีเซ็นเซอร์ที่ก้าวหน้าไป

อย่างมาก ทำให้การเรียนรู้กิจกรรมของมนุษย์มีความเป็นไปได้ที่จะกระทำการเรียนรู้โดยอัตโนมัติ อันเนื่องมาจากเซ็นเซอร์ได้มีขนาดเล็กและใช้พลังงานต่ำลง, ต้นทุนต่ำและความจุที่มากขึ้นพร้อมทั้งความก้าวหน้าในด้านเครือข่ายไร้สาย การประมวลผลข้อมูลและการเรียนรู้ด้วยเครื่อง ความก้าวหน้าเหล่านี้ได้ทำให้นักวิจัยเปลี่ยนการให้ความสำคัญจากการรวบรวมข้อมูลในระดับต่ำไปเพื่อจดจำหรือจำแนกกิจกรรมที่เกิดจากข้อมูลที่ได้รับจากเซ็นเซอร์ มาเป็นการอนุมานและการรับรู้ระดับสูงขึ้น

เนื่องจากเหมืองกระบวนการ[1][2] เป็นเทคนิคที่ใช้ในการค้นหาคุณค่า (value) จากข้อมูลจำนวนมากมายที่เกิดขึ้นจริงในระบบสารสนเทศขององค์กร หรือแม้แต่ในโทรศัพท์มือถือ ที่มีการบันทึกข้อมูลจากเซ็นเซอร์ต่างๆ เพื่อใช้วิเคราะห์กิจกรรมต่างๆ ที่เกิดขึ้นไว้ในบันทึกเหตุการณ์ (Event log) โดยมีวัตถุประสงค์หลัก 3 ประการคือ การค้นพบกระบวนการ (process discovery), การตรวจสอบความสอดคล้อง (conformance checking), และการปรับปรุงให้ดีขึ้น (enhancement) ผลลัพธ์ของการทำเหมืองกระบวนการสามารถนำเสนอได้ทั้งในรูปแบบของกราฟ, แบบจำลองของกิจกรรมที่เกิดขึ้นในระบบ หรือ ข้อมูลทางสถิติต่างๆ ทำให้ผู้ใช้สามารถมองเห็นความสัมพันธ์ระหว่างกิจกรรมต่างๆ เวลาที่ใช้ในแต่ละกิจกรรม จากข้อมูลที่ถูกบันทึกไว้ในบันทึกเหตุการณ์ได้อย่างลึกซึ้งตั้งแต่เริ่มต้นกระบวนการจนจบกระบวนการทำให้ผู้ใช้สามารถเข้าใจถึงกระบวนการที่เกิดขึ้นจริงๆ ในระบบ ไม่ใช่กระบวนการที่เกิดจากข้อมูลที่จำลองขึ้นมา การที่ผู้ใช้สามารถมองเห็นถึงปัญหาที่แท้จริง เช่น ปัญหาคอขวด เกิดขึ้น

ที่ใด เนื่องจากสาเหตุอะไร ก็จะทำให้สามารถบริหารจัดการหรือปรับปรุงให้ดีขึ้นได้อย่างมีประสิทธิภาพและประสิทธิผลต่อไป

การวิจัยนี้จึงเน้นไปที่การเตรียมข้อมูลเพื่อการทำเหมืองกระบวนการจากข้อมูลเซ็นเซอร์ที่รวบรวมมาได้จากโทรศัพท์มือถือ เนื่องจากเป็นอุปกรณ์พกพาที่ปัจจุบันคนส่วนใหญ่พกติดตัวตลอดเวลาอยู่แล้ว ทำให้ได้ข้อมูลจริงที่เกี่ยวข้องกับบุคคลนั้นๆ โดยทำการศึกษาในระดับกิจกรรม โดยใช้กรณีศึกษาจากกิจกรรม การเดิน การวิ่ง การอยู่นิ่ง ของแต่ละบุคคล เพื่อแสดงให้เห็นพฤติกรรมของบุคคล

## 2. เครื่องมือในการทำเหมืองกระบวนการ [1]

ในการเตรียมข้อมูลเพื่อใช้ในการทำเหมืองกระบวนการนั้นต้องเข้าใจเครื่องมือที่จะใช้ในการวิเคราะห์ข้อมูลด้วย เครื่องมือที่ใช้ในการทำเหมืองกระบวนการ ปัจจุบันมีเครื่องมือจำนวนมากที่สนับสนุนการทำเหมืองกระบวนการ มีทั้งที่อยู่ในรูปแบบของ open source และ ที่เป็นเชิงพาณิชย์ ในบทความนี้จะได้แนะนำตัวอย่างที่มาจากทั้งสองค่ายที่ได้รับความนิยมอย่างมากและสามารถ download นำมาทดลองใช้กันได้เลยได้แก่ โปรแกรม ProM และ Disco

- ก) ProM [3] ([www.processmining.org](http://www.processmining.org)) เป็นตัวอย่างของซอฟต์แวร์ที่ใช้ในกลุ่มของนักวิชาการหรือผู้ที่มีความรู้ความเข้าใจในหลักการการทำเหมืองกระบวนการอยู่พอสมควร เป็นซอฟต์แวร์ประเภท open source ที่มีชุดของเครื่องมือที่ใช้ในการสนับสนุนเทคนิคในการทำเหมืองกระบวนการที่หลากหลายโดย

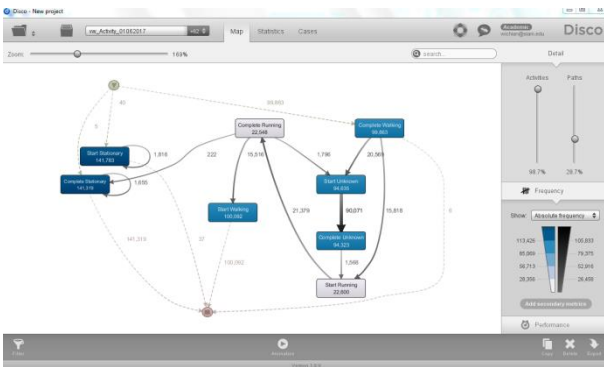
การสนับสนุนสถาปัตยกรรมแบบ Plug-in ทำให้นักวิจัยสามารถเพิ่มอัลกอริทึมใหม่ๆ หรือปรับปรุงบางส่วนเข้าไปเพิ่มเติมได้ ProM ใช้ข้อมูลเข้าในรูปแบบที่เป็นมาตรฐานของ MXML, SA-MXML, หรือ XES รุ่นล่าสุดของ ProM อยู่ที่ version 6.6 ดังแสดงในรูปที่ 1



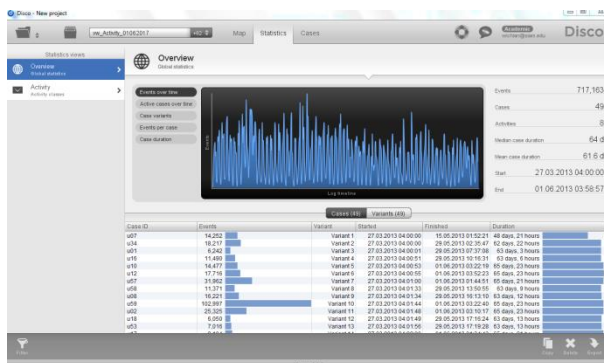
รูปที่ 1 ProM version 6.6

- ข) Disco [4] เป็นเครื่องมือเชิงพาณิชย์ที่สนับสนุนเทคนิคในการทำเหมืองกระบวนการที่มีการออกแบบส่วนเชื่อมต่อกับผู้ใช้ที่เป็นมิตรมีความสะดวกในการใช้งานและสามารถเรียนรู้ได้โดยไม่จำเป็นต้องมีความรู้และประสบการณ์ในการทำเหมืองกระบวนการมากนักแต่ก็มีข้อจำกัดอยู่ที่เป็นโปรแกรมที่สนับสนุนการทำงานหลักๆ เฉพาะอัลกอริทึม Fuzzy miner [4] เท่านั้นดังแสดงในรูปที่ 6 เป็น Disco version 1.9.9 ในส่วนของข้อมูลเข้าสามารถรับได้ในรูปแบบของ MXML, SA-MXML, หรือ XES เช่นเดียวกับ ProM แต่ที่สำคัญ และเป็นจุดเด่นของ Disco อีกประการหนึ่งคือ Disco สามารถรับข้อมูลที่อยู่

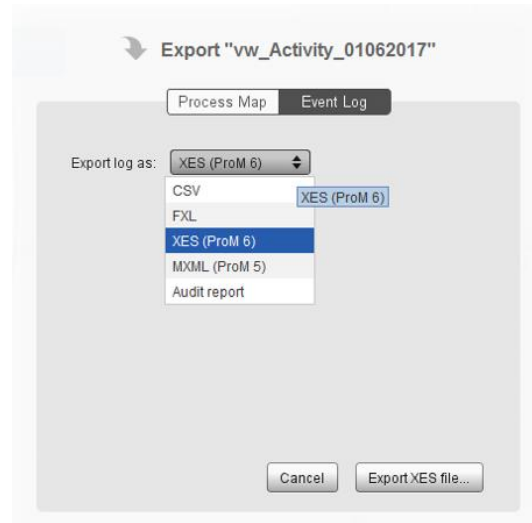
ในรูปแบบของ csv ได้ ทำให้มีความสะดวกในการใช้งาน เนื่องจากสามารถใช้ข้อมูลจาก spreadsheet ได้ พร้อมทั้งสามารถ export ข้อมูลออกไปในรูปแบบต่างๆ เพื่อนำไปใช้ในการวิเคราะห์ด้วยอัลกอริทึมอื่นๆ ที่โปรแกรม ProM สามารถทำการวิเคราะห์ได้อีกด้วยดังแสดงในรูปที่ 2 ถึงรูปที่ 4



รูปที่ 2 แบบจำลองกระบวนการ ที่แสดงในโปรแกรม Disco version 1.9.9



รูปที่ 3 ค่าทางสถิติที่แสดงในโปรแกรม Disco version 1.9.9



รูปที่ 4 รูปแบบข้อมูลที่สามารถ Export ได้จากโปรแกรม Disco version 1.9.9

### 3. บันทึกเหตุการณ์ [1]

หัวใจหลักในการทำเหมืองกระบวนการคือ บันทึกเหตุการณ์เนื่องจากเป็นข้อมูลชุดของเหตุการณ์ที่เกิดขึ้นจริงในระบบสารสนเทศที่ถูกบันทึกไว้ ในการทำเหมืองกระบวนการนั้นจะใช้บันทึกเหตุการณ์เป็นข้อมูลนำเข้าเพื่อทำการวิเคราะห์ตามเทคนิคต่างๆ ของการทำเหมืองกระบวนการโดยทั่วไปบันทึกเหตุการณ์จะประกอบไปด้วยสารสนเทศที่เกี่ยวกับเหตุการณ์ที่เกิดขึ้นที่อ้างถึงกิจกรรม (activity) ต่างๆ ที่เกิดขึ้น และ กรณี (case) โดยที่ กรณี หรือ บางครั้งเรียกว่า กระบวนการ ณ ขณะใดขณะหนึ่ง (process instance) คือ สิ่งที่เราต้องการบริหารจัดการหรือต้องการวิเคราะห์นั่นเอง เช่น ค่าสั่งซื้อของลูกค้าผู้ป่วยในโรงพยาบาล เป็นต้น ส่วนกิจกรรม หมายถึง การดำเนินการที่มีในกรณีนั้นๆ เช่น กิจกรรมของผู้ป่วยในโรงพยาบาลอาจประกอบด้วยกิจกรรมดังนี้ การลงทะเบียน การรอฟบแพทย์ พบแพทย์ จ่ายเงิน และรับยา เป็นต้น นอกจากนั้นแล้วบันทึกเหตุการณ์มักจะ

ประกอบด้วยประทับเวลา (timestamp) เพื่อเป็นการบ่งชี้เวลาที่เหตุการณ์ต่างๆ เกิดขึ้น เพื่อใช้บ่งบอกลำดับการเกิดขึ้นของกิจกรรม ในบางกิจกรรมอาจจะมีบุคคลที่กระทำหรือเข้ามาเกี่ยวข้องกับกิจกรรมนั้นๆ ในบันทึกเหตุการณ์หากมีการบันทึกบุคคลที่เป็นผู้กระทำในแต่ละกิจกรรมไว้ด้วยแล้ว ข้อมูลเหล่านี้สามารถนำไปใช้ในการทำเหมืองกระบวนการเพื่อหาเครือข่ายทางสังคม (social network) ของบุคคลากรเหล่านี้ได้ว่ามีเกี่ยวข้องกับอย่างไรบ้าง เป็นต้น ดังนั้นการสร้างบันทึกเหตุการณ์ให้อยู่ในรูปแบบที่มาสามารถนำไปใช้ได้โดยโปรแกรมการทำเหมืองกระบวนการจึงเป็นสิ่งที่จำเป็นและไม่สามารถหลีกเลี่ยงได้

#### 4. การเตรียมข้อมูลจากอุปกรณ์เซ็นเซอร์

งานวิจัยนี้ต้องการนำข้อมูลที่เก็บรวบรวมได้จากอุปกรณ์เซ็นเซอร์เพื่อนำไปใช้ในการวิเคราะห์ข้อมูลในเชิงพฤติกรรมในระดับกิจกรรมด้วยเทคนิคการทำเหมืองกระบวนการ ไม่ได้เน้นไปที่การรู้จำกิจกรรมจากข้อมูลเซ็นเซอร์ในโทรศัพท์มือถือโดยตรง เนื่องจากมีการวิจัยในเรื่องนี้เป็นจำนวนมากและได้ผลการรู้จำกิจกรรมที่ได้ผลดีเป็นที่น่าพอใจ [5] ดังนั้นในส่วนนี้จะได้กล่าวถึงการเตรียมบันทึกเหตุการณ์จากข้อมูลเซ็นเซอร์ที่ได้จากโทรศัพท์มือถือที่มีการรู้จำกิจกรรมแล้ว โดยใช้ข้อมูลจาก StudentLife Dataset [6] ซึ่งเป็นข้อมูลที่รวบรวมจากเซ็นเซอร์ของโทรศัพท์มือถือของนักศึกษาที่เข้าร่วมการวิจัย เนื่องจากข้อมูลมีเป็นจำนวนมากงานวิจัยนี้จึงเน้นเฉพาะข้อมูลกิจกรรมทางกายภาพที่สามารถตรวจจับได้จากเซ็นเซอร์ในโทรศัพท์ ที่ผ่านกระบวนการรู้จำ

กิจกรรมแล้วเพื่อใช้เป็นกรณีศึกษา โดยมีกิจกรรมที่เกิดขึ้นได้ดังนี้ อยู่นิ่งๆ (stationary), กำลังเดิน (walking), กำลังวิ่ง (running) และ ไม่ทราบ (unknown) ข้อมูลเหล่านี้มีการจัดเก็บในรูปแบบของ csv ดังแสดงในรูปที่ 5

ข้อมูลที่แสดงในรูปที่ 5 แสดงการจัดเก็บข้อมูลที่ประกอบไปด้วยข้อมูล 2 ส่วนด้วยกันคือในคอลัมน์ A เป็นเวลาประทับ (timestamp) แสดงเวลาที่เกิดขึ้นของกิจกรรม ในรูปแบบของ bigint ส่วนในคอลัมน์ B ได้ของกิจกรรมที่เกิดขึ้นดังนี้

	A	B
1	1365793104	1
2	1365793112	1
3	1365793122	1
4	1365793127	0
5	1365793323	0
6	1365793346	1
7	1365793359	0
8	1365793367	0
9	1365793581	0
10	1365793583	0
11	1365793586	0
12	1365793588	0
13	1365793593	0
14	1365793596	0
15	1365793599	0
16	1365793601	0
17	1365793604	0
18	1365793606	0
19	1365793609	0
20	1365793794	0
21	1365793805	0
22	1365793807	0
23	1365793810	1
24	1365793812	1
25	1365793815	1
26	1365793817	1
27	1365793820	1
28	1365793823	1
29	1365793825	1
30	1365793828	1
31	1365793830	1
32	1365793835	1

รูปที่ 5 ข้อมูลการเซ็นเซอร์ที่ใช้ในการตรวจจับกิจกรรมในรูปแบบของ csv

- 0 อยู่นิ่งๆ
- 1 กำลังเดิน
- 2 กำลังวิ่ง
- 3 ไม่ทราบ

โดยมีการจัดเก็บข้อมูลของนักศึกษาแต่ละคนของแต่ละเซ็นเซอร์ไว้ในคนละไฟล์แยกจากกัน ตัวอย่างเช่น ถ้ามีอุปกรณ์เซ็นเซอร์ที่ต้องการจัดเก็บข้อมูล 5 อุปกรณ์ก็จะมีไฟล์ข้อมูลในรูปแบบของ csv 5 ไฟล์สำหรับนักศึกษา 1 คน ถ้าต้องการศึกษาพฤติกรรมของนักศึกษาหลายคนจะต้องมีกระบวนการในการรวมข้อมูลเหล่านี้เข้าด้วยกัน

งานวิจัยนี้จึงนำเสนอวิธีการนำข้อมูลจากเซ็นเซอร์เหล่านี้มารวมกัน เพื่อให้ข้อมูลในรูปแบบที่สามารถนำไปประมวลผลด้วยเทคนิคการทำเหมืองกระบวนการได้ต่อไปดังนี้

#### ก) นำข้อมูลเข้าฐานข้อมูล

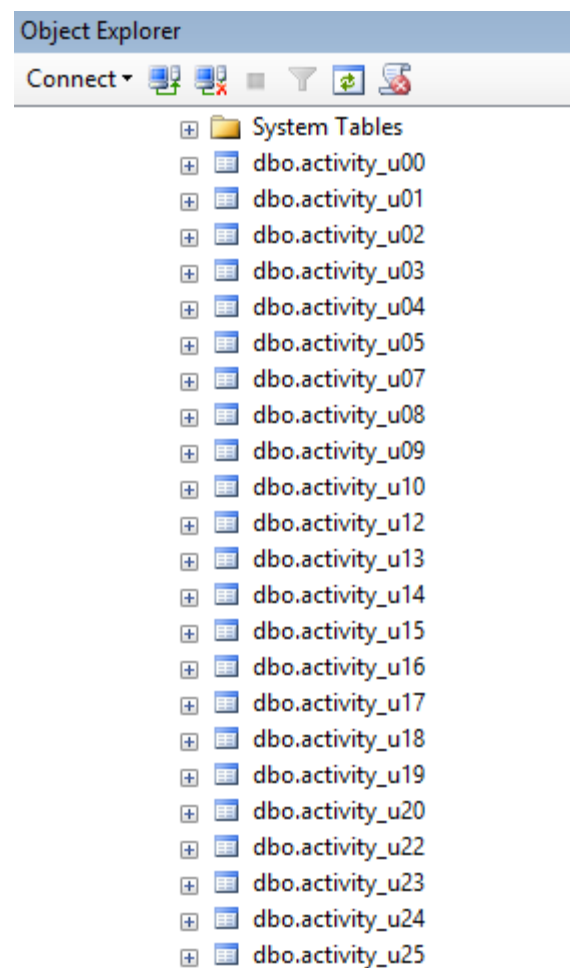
โดยทั่วไปข้อมูลที่ได้รับจากอุปกรณ์เซ็นเซอร์จะอยู่ในรูปแบบของไฟล์ข้อความ หรือ csv ทำให้ยากต่อการบริหารจัดการข้อมูล โดยเฉพาะอย่างยิ่งเมื่อมีจำนวนเซ็นเซอร์และจำนวนนักศึกษาเป็นจำนวนมากเพื่อความสะดวกในการบริหารจัดการ และการปรับเปลี่ยนมุมมองในการวิเคราะห์ที่หลากหลาย จึงได้นำข้อมูลที่อยู่ในรูปแบบ csv หนึ่งไฟล์ มาสร้างเป็นตารางในฐานข้อมูลหนึ่งตาราง ดังแสดงในรูปที่ 6

#### ข) แปลงข้อมูลและการเพิ่มคำบรรยายข้อมูล

การแปลงข้อมูลโดยเฉพาะอย่างยิ่งในข้อมูลที่เป็นประทับเวลาให้อยู่ในรูปแบบที่ใช้ได้โดยโปรแกรม Disco และ ProM

#### ค) เพิ่ม UserID

ข้อมูลที่ได้จากอุปกรณ์เซ็นเซอร์มีเฉพาะข้อมูลที่เซ็นเซอร์ตรวจจับได้ ดังนั้นในการวิเคราะห์ข้อมูลของผู้ใช้หลายคนจึงจำเป็นต้องมีการเพิ่ม UserID เข้าไปเพื่อเป็นการแยกแยะว่าเป็นข้อมูลของผู้ใดก่อนการรวมข้อมูลเข้าด้วยกัน ดังแสดงในรูปที่ 7



รูปที่ 6 ตัวอย่างการนำข้อมูลเซ็นเซอร์กิจกรรมของนักศึกษาแต่ละคนเข้าสู่ฐานข้อมูล



1	DateTimeStamp	CDateTimeStamp	ActivityInference	Description
2	1364395334	27/03/2013 14:42:14	0	Stationary
3	1364395337	27/03/2013 14:42:17	0	Stationary
4	1364395340	27/03/2013 14:42:20	0	Stationary
5	1364395342	27/03/2013 14:42:22	0	Stationary
6	1364395345	27/03/2013 14:42:25	0	Stationary
7	1364395347	27/03/2013 14:42:27	3	Unknown
8	1364395350	27/03/2013 14:42:30	3	Unknown
9	1364395353	27/03/2013 14:42:33	0	Stationary
10	1364395355	27/03/2013 14:42:35	0	Stationary
11	1364395358	27/03/2013 14:42:38	0	Stationary
12	1364395360	27/03/2013 14:42:40	0	Stationary
13	1364395363	27/03/2013 14:42:43	0	Stationary
14	1364395365	27/03/2013 14:42:45	0	Stationary
15	1364395368	27/03/2013 14:42:48	0	Stationary
16	1364395371	27/03/2013 14:42:51	0	Stationary
17	1364395373	27/03/2013 14:42:53	0	Stationary
18	1364395376	27/03/2013 14:42:56	0	Stationary
19	1364395378	27/03/2013 14:42:58	0	Stationary
20	1364395561	27/03/2013 14:46:01	1	Walking
21	1364395564	27/03/2013 14:46:04	1	Walking
22	1364395566	27/03/2013 14:46:06	1	Walking

รูปที่ 7 การแปลงข้อมูลและเพิ่มคำบรรยาย

### ง) เพิ่ม UserID

ไฟล์ข้อมูลที่ได้จากอุปกรณ์เซ็นเซอร์มีเฉพาะข้อมูลที่เซ็นเซอร์ตรวจจับได้ ไม่ได้ระบุว่าเป็นของผู้ใด ดังนั้นในการวิเคราะห์ข้อมูลของผู้ใช้หลายคนจึงจำเป็นต้องมีการเพิ่ม UserID เข้าไปเพื่อเป็นการแยกแยะว่าเป็นข้อมูลของผู้ใดก่อนการรวมข้อมูลเข้าด้วยกัน เพื่อให้สามารถวิเคราะห์ข้อมูลของหลายคนในไฟล์เดียวกัน ดังแสดงในรูปที่ 8

1	DateTimeStamp	CDateTimeStamp	ActivityInference	Description	UserID	strEvent
2	1364395334	27/03/2013 14:42:14	0	Stationary	u01	Start
3	1364395337	27/03/2013 14:42:17	0	Stationary	u01	Active
4	1364395340	27/03/2013 14:42:20	0	Stationary	u01	Active
5	1364395342	27/03/2013 14:42:22	0	Stationary	u01	Active
6	1364395345	27/03/2013 14:42:25	0	Stationary	u01	End
7	1364395347	27/03/2013 14:42:27	3	Unknown	u01	Start
8	1364395350	27/03/2013 14:42:30	3	Unknown	u01	End
9	1364395353	27/03/2013 14:42:33	0	Stationary	u01	Start
10	1364395355	27/03/2013 14:42:35	0	Stationary	u01	Active
11	1364395358	27/03/2013 14:42:38	0	Stationary	u01	Active
12	1364395360	27/03/2013 14:42:40	0	Stationary	u01	Active
13	1364395363	27/03/2013 14:42:43	0	Stationary	u01	Active
14	1364395365	27/03/2013 14:42:45	0	Stationary	u01	Active
15	1364395368	27/03/2013 14:42:48	0	Stationary	u01	Active
16	1364395371	27/03/2013 14:42:51	0	Stationary	u01	Active
17	1364395373	27/03/2013 14:42:53	0	Stationary	u01	Active
18	1364395376	27/03/2013 14:42:56	0	Stationary	u01	Active
19	1364395378	27/03/2013 14:42:58	0	Stationary	u01	End
20	1364395561	27/03/2013 14:46:01	1	Walking	u01	Start
21	1364395564	27/03/2013 14:46:04	1	Walking	u01	Active
22	1364395566	27/03/2013 14:46:06	1	Walking	u01	End

รูปที่ 8 การเพิ่ม UserID

### จ) ลดจำนวนข้อมูล

จากประทับเวลาของแต่ละกิจกรรมของข้อมูล ในรูปที่ 7 ในคอลัมน์ของประทับเวลา จะเห็นได้ว่าเซ็นเซอร์ทำการเก็บรวบรวมข้อมูลของกิจกรรมทุก 3 วินาที ทำให้มีกิจกรรมที่ซ้ำๆ กันเป็นจำนวนมากเรียงต่อกันและทำให้มีจำนวนเหตุการณ์มากเป็นล้านๆ เหตุการณ์ดังนั้นในส่วนนี้จะทำการยุบรวมเหตุการณ์ของกิจกรรมเดียวกันที่มีการกระทำอย่างต่อเนื่องกันเข้าด้วยกันให้เหลือเพียงสองกิจกรรมเท่านั้นโดยมีประทับเวลาเป็นการแสดงการเริ่มต้นและจบกิจกรรม

1	DateTimeStamp	CDateTimeStamp	ActivityInference	Description	UserID	strEvent
2	1364395334	27/03/2013 14:42:14	0	Stationary	u01	Start
3	1364395345	27/03/2013 14:42:25	0	Stationary	u01	End
4	1364395347	27/03/2013 14:42:27	3	Unknown	u01	Start
5	1364395350	27/03/2013 14:42:30	3	Unknown	u01	End
6	1364395353	27/03/2013 14:42:33	0	Stationary	u01	Start
7	1364395378	27/03/2013 14:42:58	0	Stationary	u01	End
8	1364395561	27/03/2013 14:46:01	1	Walking	u01	Start
9	1364395566	27/03/2013 14:46:06	1	Walking	u01	End

รูปที่ 9 การยุบรวมกิจกรรม

จากรูปจะเห็นได้ว่าจำนวนเหตุการณ์ที่เกิดขึ้นทั้งสิ้น 22 เหตุการณ์ในรูปที่ 8 ถูกยุบรวมเหลือเพียง 9 เหตุการณ์ในรูปที่ 9 และเหลือเพียง 5 เหตุการณ์ในรูปที่ 10

1	DateTimeStamp	CDateTimeStart	CDateTimeEnd	ActivityInference	Description	UserID
2	1364395334	27/03/2013 14:42:14	27/03/2013 14:42:25	0	Stationary	u01
3	1364395347	27/03/2013 14:42:27	27/03/2013 14:42:30	3	Unknown	u01
4	1364395353	27/03/2013 14:42:33	27/03/2013 14:42:58	0	Stationary	u01
5	1364395561	27/03/2013 14:46:01	27/03/2013 14:46:06	1	Walking	u01
6						
7						

รูปที่ 10 การยุบรวมกิจกรรมและประทับเวลา

## 5. การทดลอง

ในหัวข้อนี้จะได้ทำการทดลองการนำขั้นตอนของการประมวลผลล่วงหน้าที่น่าเสนอไปลองใช้จริงกับโปรแกรม Disco และ ProM โดยรูปที่ 11 แสดงตัวอย่างของบันทึกเหตุการณ์ที่นำเข้ามาด้วยโปรแกรม Disco โดยที่แต่ละเหตุการณ์ประกอบด้วยข้อมูลดังนี้

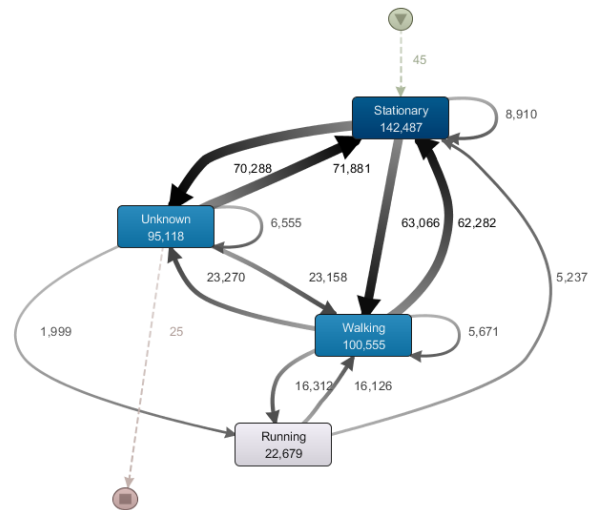
UserID	แทน กรณี
CDateTimeStart	แทน ประทับเวลาเริ่มต้น
CDateTimeEnd	แทน ประทับเวลาสิ้นสุด
Description	แทน กิจกรรม

ดังนั้นข้อมูลนำเข้าในแต่ละแถวจึงหมายถึงเหตุการณ์ของแต่ละกิจกรรมของแต่ละกรณีที่เกิดขึ้น พร้อมทั้งเวลาเริ่มต้นและสิ้นสุดของเหตุการณ์นั้น จากนั้นทำการ Export ข้อมูลออกให้อยู่ในรูปแบบของ XES เพื่อใช้เป็นข้อมูลเข้าสำหรับโปรแกรม ProM 6.6

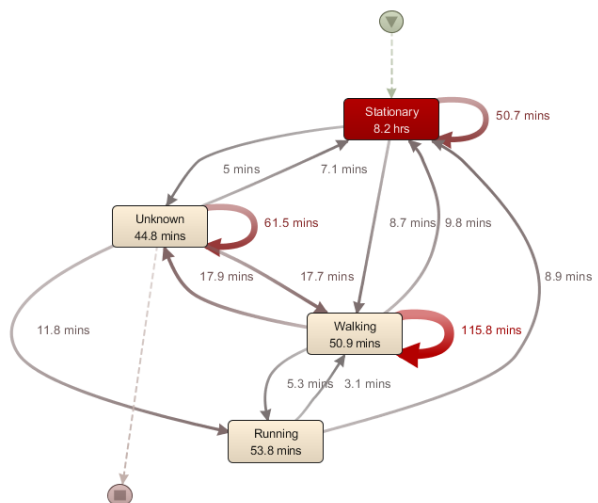
UserID	CDateTimeStart	CDateTimeEnd	ActivityInference	Description
u31	16/04/2013 17:48:08	16/04/2013 17:55:14	0	Stationary
u31	16/04/2013 17:55:17	16/04/2013 17:55:27	1	Walking
u31	16/04/2013 17:55:29	16/04/2013 17:55:32	2	Running
u31	16/04/2013 17:55:35	16/04/2013 17:55:53	1	Walking
u31	16/04/2013 17:55:55	16/04/2013 17:55:58	2	Running
u31	16/04/2013 17:55:58	16/04/2013 17:59:16	1	Walking
u31	16/04/2013 17:59:19	16/04/2013 18:00:08	0	Stationary
u31	16/04/2013 18:00:10	16/04/2013 18:03:13	1	Walking
u31	16/04/2013 18:03:13	16/04/2013 18:19:49	0	Stationary
u31	16/04/2013 18:19:52	16/04/2013 18:20:07	1	Walking
u31	16/04/2013 18:20:10	16/04/2013 18:23:13	3	Unknown
u31	16/04/2013 18:23:13	16/04/2013 18:35:53	0	Stationary
u31	16/04/2013 18:35:56	16/04/2013 18:35:59	3	Unknown
u31	16/04/2013 18:35:59	16/04/2013 18:39:35	0	Stationary
u31	16/04/2013 18:39:37	16/04/2013 18:39:40	3	Unknown
u31	16/04/2013 18:39:40	16/04/2013 18:39:43	0	Stationary
u31	16/04/2013 18:39:43	16/04/2013 18:39:58	3	Unknown
u31	16/04/2013 18:40:01	16/04/2013 18:43:19	0	Stationary
u31	16/04/2013 18:43:21	16/04/2013 18:43:32	3	Unknown
u31	16/04/2013 18:43:34	16/04/2013 18:47:27	0	Stationary
u31	16/04/2013 18:47:29	16/04/2013 18:47:32	3	Unknown
u31	16/04/2013 18:47:32	16/04/2013 18:47:47	0	Stationary
u31	16/04/2013 18:47:50	16/04/2013 18:48:06	3	Unknown
u31	16/04/2013 18:48:08	16/04/2013 19:03:20	0	Stationary
u31	16/04/2013 19:03:23	16/04/2013 19:11:32	1	Walking
u31	16/04/2013 19:11:35	16/04/2013 19:11:45	0	Stationary
u31	16/04/2013 19:11:53	16/04/2013 19:15:19	1	Walking
u31	16/04/2013 19:15:21	16/04/2013 19:19:42	0	Stationary
u31	16/04/2013 19:19:44	16/04/2013 19:19:49	3	Unknown

รูปที่ 11 ตัวอย่างของบันทึกเหตุการณ์ที่นำเข้ามาด้วยโปรแกรม Disco

หลังจากข้อมูลถูกนำเข้าแล้วก็สามารถทำการวิเคราะห์ต่างๆ ตามฟังก์ชันที่มีในโปรแกรม Disco และ ProM เช่นการวิเคราะห์ความถี่ของกิจกรรมและความหนาแน่นของการส่งต่อกิจกรรมต่างๆ หรือ การวิเคราะห์เชิงเวลาเฉลี่ยในการทำกิจกรรมต่างๆและเวลาเฉลี่ยในการทำกิจกรรมต่อไป ได้ดังแสดงในรูปที่ 12-15

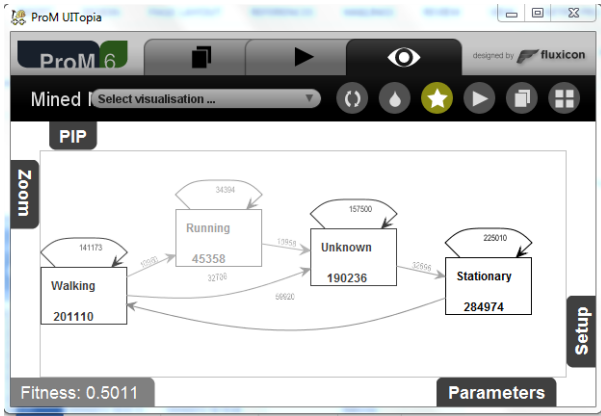


รูปที่ 12 การวิเคราะห์เชิงความถี่ด้วยโปรแกรม Disco

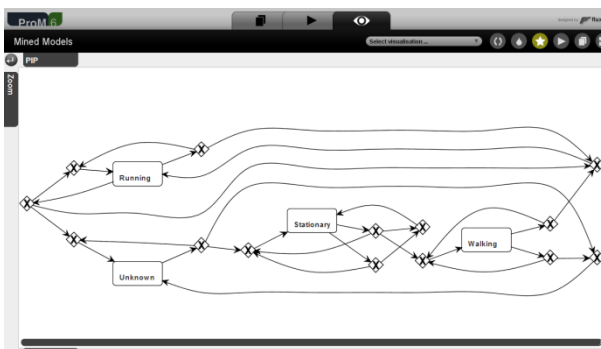


รูปที่ 13 การวิเคราะห์เชิงความเวลาคะด้วยโปรแกรม Disco





รูปที่ 14 การวิเคราะห์ Mine for a Heuristic Net Using Heuristic Miner (Annotation) ในโปรแกรม ProM



รูปที่ 15 การวิเคราะห์ Mine for a Heuristic Net Using Heuristic Miner (Semantics) ในโปรแกรม ProM

## 6. สรุปและงานในอนาคต

บทความฉบับนี้ได้นำเสนอเทคนิคการเตรียมข้อมูลจากข้อมูลที่เก็บรวบรวมได้จากเซ็นเซอร์เพื่อสร้างบันทึกเหตุการณ์เพื่อใช้เป็นข้อมูลเข้าในการทำเหมืองกระบวนการซึ่งข้อมูลเหล่านี้จะมีมากขึ้นทุกวัน การนำข้อมูลจริงที่ได้เซ็นเซอร์เพื่อนำมาวิเคราะห์เพื่อหาสาเหตุและแนวการแก้ไขจึงเป็นสิ่งที่จำเป็นและมีประโยชน์มาก วิธีการที่นำเสนอทำให้

สามารถจัดการข้อมูลที่มีอยู่อย่างกระจัดกระจายได้อย่างมีประสิทธิภาพ ตลอดจนการแก้ไขปรับปรุงข้อมูลให้สามารถวิเคราะห์ได้ในหลายมิติ

ในอนาคตจะได้มีการขยายงานวิจัยออกไปให้สามารถครอบคลุมการนำข้อมูลจากเซ็นเซอร์หลายตัวนำมารวบเข้าด้วยกันโดยผู้ใช้หลายคน และพัฒนาโครงสร้างข้อมูลที่สามารถสร้างข้อมูลออกในรูปแบบ XES ได้โดยตรงต่อไป

## เอกสารอ้างอิง

- [1] วิเชียร เปรมชัยสวัสดิ์, เหมืองกระบวนการ, วารสารวิศวกรรมศาสตร์ มหาวิทยาลัยสยาม ปีที่ 16 ฉบับที่ 1 ลำดับที่ 30 ช่วง มกราคม-มิถุนายน 2558 หน้า 1-10
- [2] W. van der Aalst. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Berlin, 2011.
- [3] TU/e Workgroup (<http://www.processmining.org>)
- [4] Fluxicon Disco (<https://fluxicon.com/disco/>)
- [5] Fu, Tun, Human Activity Recognition and Prediction, Springer, 2016.
- [6] StudentLife Dataset (<http://studentlife.cs.dartmouth.edu/dataset.html>)