

ข้อมูล ณ จุดขายเป็นข้อมูลการขายสินค้าชนิดหนึ่งทีรวบรวมได้จากซูเปอร์มาร์เก็ตหรือร้านค้าอื่นๆ โดยแต่ละรายการการเปลี่ยนแปลงของข้อมูล ณ จุดขายจะประกอบด้วย รหัสลูกค้า เวลาที่ขายสินค้า หมวดของสินค้า และราคาของสินค้า เนื่องจากข้อมูล ณ จุดขายมีข้อมูลที่มีความสัมพันธ์กับช่วงเวลา การเปลี่ยนข้อมูล ณ จุดขายให้อยู่ในรูปแบบของต้นไม้ประวัติเป็นวิธีการหนึ่งที่สามารถช่วยให้การวิเคราะห์ข้อมูลโดยการทำให้เหมือนข้อมูลมีประสิทธิภาพเพิ่มมากขึ้น

วิทยานิพนธ์ฉบับนี้นำเสนอขั้นตอนวิธี JSM ที่ถูกพัฒนาขึ้นเพื่อการทำเหมืองข้อมูลต้นไม้ประวัติ โดยขั้นตอนวิธี JSM จะรับข้อมูลเซตของต้นไม้ประวัติและค่าสนับสนุนขั้นต่ำจากผู้ใช้นั้นผลลัพธ์ของขั้นตอนวิธีคือ เซตของต้นไม้ส่วนย่อยที่ปรากฏขึ้นบ่อยครั้งที่มีค่าสนับสนุนไม่น้อยกว่าค่าสนับสนุนขั้นต่ำที่กำหนดไว้ ซึ่งในการพัฒนาขั้นตอนวิธี JSM จะใช้ขั้นตอนวิธี FREQT ซึ่งเป็นขั้นตอนวิธีหนึ่งที่ได้รับการยอมรับในการทำเหมืองข้อมูลต้นไม้ โดยการปรับปรุงขั้นตอนวิธีจะอาศัยลักษณะเฉพาะของต้นไม้ประวัติ คือ ต้นไม้ประวัติจะมีลักษณะโครงสร้างที่เหมือนกันและตำแหน่งของคำอธิบายของช่วงเวลาต่างๆ จะมีตำแหน่งที่แน่นอน นอกจากนี้ขั้นตอนวิธี JSM ยังได้ประยุกต์ใช้คุณสมบัติของ Apriori เข้าไปเพื่อเพิ่มประสิทธิภาพการทำงานอีกด้วย ผลที่ได้จากการปรับปรุงขั้นตอนวิธีด้วยวิธีการดังที่ได้กล่าวมาแล้ว เมื่อพิจารณาความซับซ้อนเชิงคำนวณของขั้นตอนวิธี JSM คือ $O(k \times \text{Sum } F_i \times \text{Sum } R \times N)$ ในขณะที่สำหรับขั้นตอนวิธี FREQT เมื่อนำมาประยุกต์ใช้ในการทำเหมืองข้อมูลต้นไม้ประวัติโดยตรงขั้นตอนวิธี FREQT จะมีความซับซ้อนเชิงคำนวณคือ $O(16^m \times \text{Sum } F_i \times \text{Sum } R \times N)$

ในส่วนของการทดลองนั้นจะทำการทดสอบประสิทธิภาพในการทำงานของขั้นตอนวิธีทั้งสองใน 2 กรณีคือ กรณีที่จำนวนต้นไม้ส่วนย่อยที่ปรากฏขึ้นบ่อยครั้งมีจำนวนที่ค้นหาได้เพิ่มขึ้นและกรณีที่จำนวนต้นไม้ประวัติมีจำนวนเพิ่มขึ้น ซึ่งจากผลการทดลองการทำเหมืองข้อมูลต้นไม้ประวัติและจากการคำนวณความซับซ้อนเชิงคำนวณของทั้งสองขั้นตอนวิธี สามารถสรุปได้ว่าขั้นตอนวิธี JSM ที่พัฒนาขึ้นสามารถทำเหมืองข้อมูลต้นไม้ประวัติได้มีประสิทธิภาพมากกว่าขั้นตอนวิธี FREQT

Point of Sale (POS) data is a kind of sale data collected from supermarkets or similar types of stores. Typically, each transaction of the POS data composes of user ID, date of sale, product category, and also price. As POS data contain timing information, Historical Tree (HT) is one way to represent the POS data. Since the representation can help improving effectiveness when data analysis, i.e. data mining, is to be processed.

In this thesis, an algorithm, named JSM, for HT mining is proposed. Such the algorithm takes a set of HT, and the minimum support threshold from the users, subsequently, it outputs the set of frequent subtrees according to the threshold. The proposed algorithm is developed based on a well-known tree mining algorithm, FREQT algorithm. The improvement is based on a few specific properties of the HTs that they have fixed structure and fixed-location label. Also, the Apriori property has embedded to improve the proposed algorithm. All the improvements results that the complexity of the JSM algorithm is $O(k \times Sum_F_k \times Sum_R \times N)$, while applying the FREQT algorithm directly has $O(16^m \times Sum_F_k \times Sum_R \times N)$ complexity.

The proposed algorithm is validated by thorough experiments. There are two scenarios which the algorithm is investigated, i.e. when the number of frequent subtrees is increasing, and when the number of historical trees is increasing. From experiment result and the complexity of both algorithms, JSM algorithm can mine historical tree be efficiency than FREQT algorithm.