

ในยุคซึ่งข้อมูลถูกสร้างขึ้นมามากมายมหาศาล การเผยแพร่ฐานข้อมูลส่วนบุคคลสำหรับการทำเหมืองข้อมูลจำเป็นที่จะต้องมีการรักษาความเป็นส่วนตัว เนื่องจากฐานข้อมูลที่เผยแพร่สามารถถูกระบุตัวตนได้อีกครั้งโดยการเชื่อมต่อกับฐานข้อมูลที่เผยแพร่อื่นๆ ดังนั้นจึงจำเป็นที่จะต้องทำการแปลงฐานข้อมูลที่เผยแพร่ให้มีการรักษาความเป็นส่วนตัว

ในการแปลงฐานข้อมูลนี้จะต้องมีการคำนึงถึงคุณภาพข้อมูลสำหรับการทำเหมืองข้อมูลด้วย กล่าวคือ ผลกระทบต่อคุณภาพข้อมูลที่เกี่ยวข้องกับการทำเหมืองข้อมูลจะต้องน้อยที่สุด อย่างไรก็ตามปัญหาการแปลงข้อมูลให้มีการรักษาความเป็นส่วนตัวและมีผลกระทบต่อคุณภาพข้อมูลน้อยที่สุดถูกพิสูจน์แล้วว่าคือ ปัญหาที่เป็นเอ็นพีแบบยาก ยิ่งไปกว่านั้นกระบวนการจำแนกแต่ละกระบวนการมีความแตกต่างกันในการค้นหาความรู้ในการทำเหมืองข้อมูล ดังนั้นได้นำเสนอตัววัดคุณภาพข้อมูลสำหรับการจำแนก ซึ่งเป็นกระบวนการที่สำคัญสำหรับการทำเหมืองข้อมูล

ในงานวิจัยนี้ได้นำเทคนิค  $k$ -Anonymity มาใช้ในการพิจารณาในกระบวนการแปลงข้อมูลให้มีการรักษาความเป็นส่วนตัว และทำการรักษาคุณภาพข้อมูลของฐานข้อมูลที่ถูกแปลงสำหรับการจำแนกแบบความสัมพันธ์ โดยทำการสร้างตัววัดคุณภาพข้อมูลสำหรับการจำแนกแบบความสัมพันธ์ขึ้นมาใหม่ ที่ชื่อว่า Frequency-based Classification

ตัววัดคุณภาพข้อมูลจะถูกพิสูจน์โดยการทดลอง ซึ่งแสดงให้เห็นว่าตัววัดคุณภาพข้อมูลมีประสิทธิภาพสำหรับการจำแนกแบบความสัมพันธ์

การพัฒนาขั้นตอนวิธีแบบศึกษาสำนึกสำหรับการรักษาความเป็นส่วนตัวของข้อมูลและคุณภาพข้อมูล นำเสนอด้วยสองขั้นตอนวิธีคือ MHT และ MCCRT และตรวจสอบขั้นตอนวิธีด้วยจากการทดลองพบว่า MCCRT ไม่เพียงแต่มีประสิทธิภาพและยังมีประสิทธิภาพสูงอีกด้วย

Since individual data are being collected everywhere in the era of data explosion, Privacy preservation is an essential issue in the release of data for mining purposes. The released data can be re-identified by being joined with other public databases. Thus, data transformation to ensure privacy preservation is needed.

Meanwhile, the transformed data must have quality for the intended data mining task, i.e. the impact on the data quality with regard to the data mining task must be minimized. However, the data transformation problem to preserve the data privacy while minimizing the impact has been proven as an NP-hard. Also, for classification mining, each of the classification approach may use different approaches to discover knowledge. In such a context, the measure of the data quality was proposed in terms of the classification, which is one of the most important tasks in data mining.

In this research, the k-Anonymity technique is considered in case of the transformation approach for preserving data privacy. The data quality is maintained in the scenarios that are the transformed data is used to build the associative classification models. A new associative classification data quality measure, frequency-based data quality measure was proposed.

Consequently, the proposed measure was proved by experiments. The experiment result has shown that our proposed measures have effectively reflect that the metrics of data quality for the associative classification problem.

The heuristic approach was proposed to preserve the privacy and maintain the data quality in two algorithms is the Maximum Height Transformation algorithm (MHT) and the Minimum Classification Correction Rate Transformation algorithm (MCCRT). Consequently, the proposed approaches were validated by experiments. From the experiments, the MCCRT is not only efficient, but also highly effective.