บทคัดย่อ

โครงการวิจัยนี้เกี่ยวกับการพัฒนาขั้นตอนวิธี (algorithm) การทำเหมืองข้อมูลเพื่อหาองค์ความรู้ใหม่จากชุดข้อมูลหรือฐานข้อมูล ขั้นตอนวิธีนี้ถูกพัฒนาขึ้นเพื่อใช้สำหรับข้อมูลความปลอดภัยในงานอุตสาหกรรม ซึ่งได้รับจากผู้ใช้ภาคอุตสาหกรรมของโครงการวิจัย ข้อมูลเหล่านี้ถูกรวบรวมจากอุบัติเหตุที่เกิดขึ้นในสถานที่ทำงานภาคอุตสาหกรรม ข้อมูลความปลอดภัยประกอบด้วย 1) ข้อมูลเกี่ยวกับพนักงานที่ได้รับบาดเจ็บ (หรือที่เกี่ยวข้องกับอุบัติเหตุ) ได้แก่ อายุ เพศ ประสบการณ์การทำงาน เป็นต้น 2) ข้อมูลเกี่ยวกับการเกิดอุบัติเหตุ เช่น วัน เวลา สถานที่ และ 3) ข้อมูลเกี่ยวกับผลที่เกิดขึ้น ได้แก่ ระดับความรุนแรงของอุบัติเหตุ ประเภทของการบาดเจ็บ และค่าใช้จ่ายที่เกิดขึ้น ขั้นตอนวิธีนี้ใช้หลักความคิด ของเซ็ตเบื้องต้น (elementary set concept) และเทคนิคทางฐานข้อมูลในการสร้างความสัมพันธ์ระหว่างกลุ่มตัวแปรที่เกี่ยวกับอุบัติเหตุและพนักงานและกลุ่มตัวแปรเกี่ยวกับผลที่เกิดจากอุบัติเหตุ ความสัมพันธ์ที่ได้ถูกแสดงในรูปของกฎการเชื่องโยงแบบ ถ้า-แล้ว (IF-THEN association rule) โดยส่วนเงื่อนไขของกฎ (IF statement) ประกอบด้วยชุดตัวแปรและค่าของชุดตัวแปรที่เป็นเงื่อนไข (โดยส่วนใหญ่เกี่ยวข้องกับอุบัติเหตุและตัวพนักงาน) และส่วนผลลัพธ์ของกฎ (THEN statement) ประกอบด้วยตัวแปรและค่าของตัวแปร (หรือชุดตัวแปร) ที่เป็นผลลัพธ์ของอุบัติเหตุ โดยตอนแรกขั้นตอนวิธีที่พัฒนาขึ้นจะสร้างกฎทั้งหมดที่เป็นไปได้ตามรูปแบบของกฎที่กำหนดโดยผู้ใช้ (เช่น กฎแบบหนึ่งต่อหนึ่ง กฎแบบสองต่อหนึ่ง) จากนั้นขั้นตอนวิธีจะทำการกรองกฎที่ไม่สำคัญออกโดยใช้วิธีทดสอบทางสถิติและตัวกรองอื่นๆ กฎที่ผ่านตัวกรองและผ่านการทดสอบทางสถิติจะถูกรายงานในผลการทำเหมืองข้อมูล เมื่อการพัฒนาแล้วเสร็จ ขั้นตอนวิธีได้ถูกทดลองใช้กับข้อมูลจริงจากภาคอุตสาหกรรมเพื่อทดสอบประสิทธิผล นอกจากนี้ ยังทำการเปรียบเทียบสมรรถนะและประสิทธิผลกับขั้นตอนวิธีการทำเหมืองข้อมูลมาตรฐานวิธีหนึ่ง พบว่าสมรรถนะและผลลัพธ์ที่ได้เป็นที่น่าพอใจ ตัวอย่างการนำไปใช้ ผลลัพธ์และคำอธิบายกฎสำคัญที่พบได้ถูกแสดงในรายงานฉบับนี้ด้วย

คำสำคัญ: การทำเหมืองข้อมูล; การหาองค์ความรู้; ข้อมูลความปลอดภัยในงานอุตสาหกรรม; กฎการเชื่อมโยง

ABSTRACT

This research project involved the development of a data mining algorithm for knowledge discovery from a dataset or database. The algorithm was constructed for the application of industrial safety. Industrial safety data, provided by the industrial user of the project, were collected from many records of accidents that occurred in industrial workplaces. The data included information related to 1) injured workers, e.g. age, gender, work experience, 2) accident occurrences, such as date and time, place; and 3) accident outcomes, i.e., severity level of accidents, types of injury, and costs incurred. The algorithm implemented elementary set concept and data base manipulation function to generate useful relationships between accident and worker related attributes and accident outcomes. The relationships obtained were in the form of IF-THEN association rules, where the IF statement contained set of condition attributes and their values (mostly related to worker and accident occurrence) and the THEN statement included attribute(s) and values that represented accident decision outcome(s). First, the algorithm would generate all possible rules according to a user-specified rule format (e.g. 1-1 rule, 2-1 rule). Then, the algorithm would apply some user-specified filters and a statistical test to identify important rules. The rules that passed the significance test and filters were then reported in the solution. After developed, the algorithm was extensively tested on a set of actual industrial safety data to verify its effectiveness. In addition, the algorithm performance and effectiveness were compared with a standard association rule algorithm. The performance and mining results were satisfactory. A numerical example to demonstrate the algorithm usage in mining the safety data was provided, along with example of reported significant rules and their interpretations.

Keywords: Data mining; knowledge discovery; industrial safety data; association rule