

บทคัดย่อภาษาไทย

ในระบบสารสนเทศศาสตร์สุขภาพ ข้อมูลทุกด้านที่เกี่ยวข้องกับสุขภาพของประชาชนจะถูกบันทึกลงฐานข้อมูลเฉพาะท้องถิ่นและเก็บรักษาไว้ที่หน่วยงานสาธารณสุขแต่ละแห่ง ฐานข้อมูลแบบกระจายเหล่านี้มีความรู้ที่เป็นประโยชน์ซ่อนอยู่ ความรู้ในลักษณะของการค้นหาความสัมพันธ์ระหว่างแอททริบิวต์หรือลักษณะประจำภายในฐานข้อมูล สามารถเรียนรู้ได้ด้วยการทำเหมืองความสัมพันธ์ แต่การเรียนรู้แบบนี้มักจะมีข้อกำหนดเบื้องต้นว่าข้อมูลจะต้องถูกรวบรวมไว้เป็นไฟล์เดียว ซึ่งในระบบสารสนเทศศาสตร์สุขภาพที่ใช้งานอยู่จริง ฐานข้อมูลแต่ละแห่งจะมีข้อมูลในปริมาณมาก เมื่อรวมข้อมูลในทุกฐานข้อมูลเป็นไฟล์ขนาดใหญ่ไฟล์เดียวจะไม่สามารถประมวลผลได้เนื่องจากไฟล์จะมีขนาดใหญ่เกินกว่าความสามารถของหน่วยประมวลผลโดยทั่วไปจะทำงานได้

งานวิจัยนี้ได้เสนอวิธีแก้ปัญหาดังกล่าวด้วยการหากฎความสัมพันธ์แบบกระจาย แต่การหากฎความสัมพันธ์เพื่อให้ได้แหล่งความรู้เพียงแหล่งเดียวในลักษณะนี้ทำได้ยาก เนื่องจากขั้นตอนการรวมกฎความสัมพันธ์นั้นอาจทำให้ได้กฎความสัมพันธ์ที่ขัดแย้งกันเอง หรือได้จำนวนกฎความสัมพันธ์ที่มากเกินไป หรือเกิดการขาดไปของกฎความสัมพันธ์ที่สำคัญ ดังนั้น งานวิจัยนี้ได้เสนอแนวทางแก้ไขปัญหาการหากฎความสัมพันธ์แบบกระจาย โดยในขั้นตอนการรวมกฎความสัมพันธ์จะนำมาเฉพาะกฎความสัมพันธ์ที่ปรากฏขึ้นบ่อยในทุก ๆ แหล่งความรู้ที่เรียนรู้ได้จากฐานข้อมูลย่อย จากนั้นนำกฎความสัมพันธ์ที่เรียนรู้ได้จากแต่ละฐานข้อมูลไปตรวจสอบความขัดแย้งและในขั้นตอนนี้สามารถสร้างกฎความสัมพันธ์ใหม่จากกฎความสัมพันธ์เดิมที่มีอยู่ด้วยวิธีการอนุมานเชิงตรรกศาสตร์ ซึ่งสามารถเพิ่มเติมในส่วนของกฎความสัมพันธ์ที่ขาดหายไปได้ สุดท้ายจะได้กฎความสัมพันธ์ที่มีประสิทธิภาพเพียงพอสำหรับการนำไปทำนายผลข้อมูลในอนาคตและไม่เกิดความขัดแย้งกันเอง

บทคัดย่อภาษาอังกฤษ

In health informatics, all data related to people's health are recorded in the local database and stored at each healthcare organization. These distributed databases have valuable hidden knowledge. A specific kind of knowledge revealing relations among data attributes can be extracted by means of association mining. But the normal assumption of such knowledge mining is that all data to be learned have to be collected into a single file. However, in reality each database in health informatics stores a lot of information. Therefore, gathering all tremendous data as a single source will result in a too big file size to be practically processed by a typical processing unit.

This research proposes a distributed learning method to solve mining association knowledge from distributed databases. The difficulties of such distributed learning method are that the process of combining association rules may lead to inconsistent rules, there may be too many number of association rules, and some significant association rules are possibly missing. We thus propose the distributed association rule mining method. In the combining process, association rules that appear frequently in all knowledge bases are combined and then checked for inconsistency of rules. This process can generate new association rules from original association rule set with the inference feature of first-order logic. Finally, the efficient and consistent association rules are obtained.