

บัลลังก์ นิยมศักดิ์ 2553: การจำแนกซับแฟมมีลี้ของเอนไซม์โดยใช้กฎความสัมพันธ์ที่มี
ลำดับเหตุการณ์ ปัญญาวิศวกรรมศาสตรมหาบัณฑิต (วิศวกรรมคอมพิวเตอร์) สาขา
วิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก: รองศาสตราจารย์กฤษณะ ไวยมัย, Ph.D. 69 หน้า

เทคนิคจำแนกซับแฟมมีลี้ของเอนไซม์เป็นงานวิจัยด้านชีวสารสนเทศศาสตร์ที่มีอย่าง
แพร่หลายในปัจจุบัน วิธีส่วนใหญ่ใช้ความรู้พื้นฐานทางสถิติซึ่งให้ความถูกต้องได้ดีในระดับหนึ่ง
อย่างไรก็ตาม อย่างไรก็ตามวิธีเหล่านี้ไม่ได้สนใจในเรื่องการสร้างโมเดลในการทำนายที่สามารถ
อธิบายได้และลำดับความสัมพันธ์ของแอททริบิวต์ ว่ามีผลอย่างไรต่อการสร้างโมเดลในการทำนาย

ในงานวิจัยนี้จึงได้เสนอเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบมี
ลำดับในการจำแนกซับแฟมมีลี้ของเอนไซม์ (SAC : Sequential Associative Classification for
Enzyme Subfamily Prediction) โดยรวมเทคนิคการหาความสัมพันธ์แบบมีลำดับเหตุการณ์
(Sequential Pattern Mining) และเทคนิคการจัดกลุ่มข้อมูลโดยใช้กฎความสัมพันธ์ (Associative
Classification) เข้าด้วยกันเพื่อการทำนาย โดยใช้ความรู้พื้นฐานเกี่ยวกับ “โมทีฟเชิงปฏิกิริยา
ชีวเคมี” มารวมเข้าด้วยกันกับขั้นตอนการหาความสัมพันธ์แบบมีลำดับเหตุการณ์ นอกจากนี้
ในการจัดกลุ่มข้อมูลโดยใช้กฎความสัมพันธ์ได้ถูกขยายขอบเขตออกไปโดยการพิจารณาลำดับ
ความสัมพันธ์ของแอททริบิวต์ในขั้นตอนการสร้างโมเดลในการทำนาย จากผลการทดลอง
นอกจากจะได้โมเดลจำแนกประเภทที่ง่ายต่อการตีความแล้วเทคนิค SAC นี้สามารถสรุปให้เห็น
ได้ถึงความสัมพันธ์ระหว่างลำดับแอททริบิวต์ ในสายโปรตีนมีผลต่อความแม่นยำในการทำนาย
โดยให้ความแม่นยำที่ดีที่สุดอยู่ที่ 70.95% โดยใช้ชุดข้อมูลโปรตีนเอนไซม์จากฐานข้อมูล SWISS-
PROT