

พัทริกา ณ พิกุล 2554: เทคนิคการเรียนรู้แบบกึ่งมีผู้สอนสำหรับการจำแนกฟังก์ชันของ
โปรตีน ปรินญาวิศวกรรมศาสตรมหาบัณฑิต (วิศวกรรมคอมพิวเตอร์) สาขาวิศวกรรม
คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก:
รองศาสตราจารย์กฤษณะ ไวยมัย, Ph.D. 52 หน้า

ปัญหาการทำนายฟังก์ชันการทำงานของโปรตีนเป็นปัญหาหลักทางด้านชีวสารสนเทศ
ศาสตร์(Bioinformatics) ที่มีการค้นคว้าอย่างกว้างขวาง การสร้างตัวทำนายฟังก์ชันสำหรับข้อมูล
อนุกรมโปรตีน (Protein sequence) ให้สามารถทำนายคลาสได้อย่างมีประสิทธิภาพนั้น จะต้องอาศัย
ข้อมูลที่ทราบคลาส (Labeled Data) เป็นจำนวนมาก ซึ่งข้อมูลประเภทนี้มีอยู่อย่างจำกัด ในขณะที่
ข้อมูลส่วนใหญ่ที่มี เป็นข้อมูลที่เราไม่ทราบคลาส (Unlabeled Data) ทำให้ได้ผลการทำนายที่ไม่ดีพอ
เนื่องจากข้อมูลที่ทราบคลาสนั้นมีจำนวนน้อยเกินไป

งานวิจัยนี้นำเทคนิคการเรียนรู้แบบกึ่งมีผู้สอน (Semi-Supervised Learning) มาแก้ปัญหาคณิ
ข้อมูลที่ทราบคลาสมีจำนวนน้อย โดยนำข้อมูลที่ทราบและไม่ทราบคลาสมาทำการฝึกสอนร่วมกัน
ในงานวิจัยนี้เสนอเกณฑ์การเลือกข้อมูลที่ใช้ค่าการเทียบเคียงลำดับอนุกรมโปรตีนและ
สัมประสิทธิ์แจคาร์ดมาเป็นเกณฑ์พิจารณา ในส่วนของเกณฑ์การหยุดสอนนั้น จะพิจารณาจากอัตรา
การเปลี่ยนแปลงของค่าสัมประสิทธิ์สหสัมพันธ์กำลังสอง ระหว่างข้อมูลในแต่ละคลาสและข้อมูลที่
ถูกเลือกในแต่ละรอบของการฝึกสอน ผลการทดลองบนฐานข้อมูล UniProtKB/Swiss-Prot ที่มีข้อมูล
ที่ทราบคลาสจำนวน 17,407 ตัวอย่าง และข้อมูลที่ไม่ทราบคลาสจำนวน 47,619 ตัวอย่าง และวัดผลตัว
จำแนกด้วยค่าความแม่นยำ, ค่าความเที่ยง, รีคอล และ เอฟเมเชอร์ จากผลการทดลองพบว่าค่าความแม่นยำ
เมื่อเทียบกับเทคนิคที่ใช้เกณฑ์การเลือกข้อมูลแจคาร์ดและเกณฑ์การหยุดค่าเฉลี่ยกำลังสองต่ำสุด
เพิ่มขึ้น 0.3% ค่าความเที่ยงเพิ่มขึ้น 1.00% ค่ารีคอลดีขึ้น 2.83% และค่าเอฟเมเชอร์ดีขึ้น 1.91%

ลายมือชื่อนิสิต

ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก