

ชื่อเรื่องวิทยานิพนธ์ ฟังก์ชันเคอร์เนลสำหรับการจำแนกชนิดของข้อมูลสายลำดับ
โดยวิธีซัพพอร์ตเวกเตอร์แมชชีน กรณีศึกษาการจำแนกชนิด
ฮิวแมน ลิวโคไซต์ แอนติเจน

ผู้เขียน นายภาณุวัฒน์ เมฆะ

ปริญญา วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

อาจารย์ที่ปรึกษาวิทยานิพนธ์ รองศาสตราจารย์ ดร.จิรยุทธ ไชยจารุณิช

บทคัดย่อ

วิทยานิพนธ์นี้มีวัตถุประสงค์เพื่อนำเสนอฟังก์ชันเคอร์เนลสำหรับการจำแนกชนิดของข้อมูลสายลำดับ โดยวิธีซัพพอร์ตเวกเตอร์แมชชีน ซึ่งวิธีการที่นำเสนอนี้เป็นการนำสเปกตรัมเคอร์เนลเพื่อจำแนกชนิดข้อมูลสายลำดับ โดยใช้คุณสมบัติของเวกเตอร์เข้ามาช่วยในการเรียนรู้ด้วยเครื่องแบบมีการสอน

การดำเนินการของวิธีการที่นำเสนอได้ถูกแสดงโดยใช้กรณีศึกษาในการจำแนกชนิดฮิวแมน ลิวโคไซต์ แอนติเจน (เฮชแอลเอ) ซึ่งข้อมูลของเฮชแอลเอนี้แบ่งเป็น 2 ชนิดคือ เฮชแอลเอกลุ่มที่ I และเฮชแอลเอกลุ่มที่ II โดยในแต่ละกลุ่มของเฮชแอลเอนั้นจะมีการแบ่งกลุ่มย่อยของข้อมูลลงไปอีก ซึ่งวิธีการนี้จะนำไปสู่แบบจำลองของการทำนายชนิดของข้อมูลสายลำดับประเภทนี้ต่อไป

จากการทดลองพบว่า วิธีการที่นำเสนอสามารถจำแนกชนิดข้อมูลสายลำดับของเฮชแอลเอ ซึ่งสามารถใช้กับข้อมูลสายลำดับได้โดยตรง โดยข้อมูลสายลำดับที่นำมาทดสอบในการทดลองเท่ากับ 4,403 สายลำดับ โดยทำการวัดความถูกต้องในการทำนายชนิดของข้อมูลและทำการเปรียบเทียบกับเทคนิคทันสมัยอื่นๆ

Thesis Title	Kernel Function for Classification of Sequential Data Using Support Vector Machines: A Case Study of Human Leukocyte Antigen Classification
Author	Mr. Panuwat Mekha
Degree	Master of Science (Computer Science)
Thesis Advisor	Assoc. Prof. Dr. Jeerayut Chaijaruwanich

Abstract

The objectives of this thesis are to propose a kernel function for classification of sequential data using support vector machines. The proposed method bases on spectrum kernels to classify sequences and supervised learning using vectors as features.

Our proposed method is illustrated through a case study of Human Leukocyte Antigen (HLA) classification which plays an important role in the control of self-recognition, including defense against microorganisms. The classifications of HLA molecules into HLA-I and HLA-II major classes and subclasses have been investigated.

In this work, we directly classify the HLA molecules using their primary sequences. Our method was tested on 4,403 HLA sequences. Using a ten-fold cross-validation, prediction accuracy for classifying the HLA major classes was shown very high comparing with other classification methods.