

พงศกร ชีร์รัมย์ : วิธีการหาค่า k ที่เหมาะสมในการจำแนกแบบเคเนียร์เรสเนเบอร์กับ
ข้อมูลทางการแพทย์ (THE METHODOLOGY TO FIND APPROPRIATE k FOR K-
NEAREST NEIGHBOR CLASSIFICATION WITH MEDICAL DATASETS)

อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.นิตยา เกิดประสพ, 72 หน้า

งานวิจัยนี้ได้ศึกษาปัญหาการจำแนกข้อมูลด้วยอัลกอริทึมเคเนียร์เรสเนเบอร์ ซึ่งในการ
จำแนกข้อมูลจำเป็นต้องกำหนดค่า k ซึ่งหมายถึงจำนวนข้อมูลใกล้เคียง โดยค่า k ที่กำหนดจะมีผล
ต่อประสิทธิภาพในการวิเคราะห์เพื่อจำแนกประเภทของข้อมูล หากกำหนดค่า k ไม่เหมาะสมจะ
ทำให้ค่าความแม่นยำจากการจำแนกประเภทต่ำกว่าที่ควรจะเป็น นอกจากนี้หากกำหนดค่า k มาก
เกินไปจะส่งผลให้การประมวลผลช้า และอาจทำให้ความแม่นยำลดลง การใช้วิธีปรับค่า k ไป
ตามลำดับจนกระทั่งได้ค่าความแม่นยำที่สูง ไม่เป็นที่นิยมนักเนื่องจากจะใช้เวลาในการประมวลผล
มากและอาจจะไม่ได้ค่าความแม่นยำสูงตามที่ต้องการ ดังนั้นงานวิจัยนี้ได้แนะนำวิธีการกำหนดค่า k
ที่เหมาะสมในการจำแนกข้อมูลทางการแพทย์ที่จะส่งผลให้การจำแนกข้อมูลมีค่าความแม่นยำสูง และ
เป็นการเพิ่มทางเลือกในการตัดสินใจให้กับผู้ใช้

การวิเคราะห์เพื่อให้ได้ซึ่งค่า k ที่เหมาะสมสำหรับจำแนกประเภทข้อมูลด้วยอัลกอริทึมเค
เนียร์เรสเนเบอร์ จำเป็นต้องรู้ถึงลักษณะของข้อมูล ซึ่งในงานวิจัยนี้จะมีการพิจารณาค่าทางสถิติ
เกี่ยวกับการกระจายของข้อมูล สำหรับใช้ในการวิเคราะห์เพื่อแนะนำการเลือกค่า k เพื่อเป็น
พารามิเตอร์สำคัญของอัลกอริทึมเคเนียร์เรสเนเบอร์

สาขาวิชา วิศวกรรมคอมพิวเตอร์

ปีการศึกษา 2558

ลายมือชื่อนักศึกษา _____

ลายมือชื่ออาจารย์ที่ปรึกษา _____

PONGSAKORN TEERARASSAMEE : THE METHODOLOGY TO FIND
APPROPRIATE K FOR K-NEAREST NEIGHBOR CLASSIFICATION
WITH MEDICAL DATASETS. THESIS ADVISOR : ASSOC. PROF.
NITTAYA KERDPRASOP, Ph.D., 72 PP.

K-NEAREST NEIGHBOR/ CLASSIFICATION/ MEDICAL DATA

In this research, we have studied the problem of data classification with the k-nearest neighbor (kNN) algorithm. On classifying data, the value of k, which means the number of closest data, has to be specified. The choice of k has great impact to the efficiency of data classification. The bad choice of k value results in a low classification accuracy. The too high value of k can also slow down the computation time and may decrease the accuracy. Sequentially adjusting k value until reaching the optimal one is not a practical method because it takes much processing time and the best accuracy cannot be guaranteed. Therefore, this research has suggested a way to configure the appropriate value of k based on data characteristics that can result in accurate classification. The proposed method can help users estimate appropriate value of k in a quick time.

On analyzing the k value suitable for kNN algorithm, it is necessary to know the data characteristics. In this research, we consider the distribution of data for analyzing and recommending the k value to be an important parameter for the kNN algorithm.

School of Computer Engineering

Academic Year 2015

Student's Signature _____

Advisor's Signature _____