วิทยาการวิจัยและวิทยาการปัญญา ปีที่ 8 ฉบับที่ 2, ตุลาคม 2553 – มีนาคม 2554
Research Methodology & Cognitive Science, Vol. 8, No. 2, October 2010 – March 2011

การคัดเลือกคุณลักษณะสำหรับข้อมูลที่มีจำนวนมิติมากในการจำแนกประเภท

นพมาศ อัครจันทโชติ
มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ

บทคัดย่อ

บทความนี้นำเสนอแนวทางการคัดเลือกคุณลักษณะสำหรับข้อมูลที่มีจำนวนมิติมาก การศึกษา ในงานการจำแนกประเภท การคัดเลือกคุณลักษณะเรื่องหัวข้อที่ได้รับความสนใจ และมีความจำเป็นต่อการวิเคราะห์ข้อมูลในหลายสาขาวิชา เนื่องจากความเจริญทางด้านเทคโนโลยีทำให้สามารถผลิตหรือรวบรวมข้อมูลที่มีขนาดใหญ่ แต่ในข้อมูลขนาดใหญ่มีเพียงบางคุณลักษณะหรือบางตัวแปรเท่านั้นที่มีความสำคัญต่อการทำนายการเป็นสมาชิกกลุ่ม จึงมีความจำเป็นที่จะต้องมีการคัดเลือกเฉพาะคุณลักษณะที่มีความสำคัญเหล่านั้น บทความนี้ได้สรุปแนวทางการคัดเลือกคุณลักษณะ รวมทั้งเปรียบเทียบข้อดีและข้อเสียของแต่ละแนวทาง

คำสำคัญ: การจำแนกประเภท/ การคัดเลือกคุณลักษณะ/ ข้อมูลที่มีจำนวนมิติมาก

Feature Selection for High-dimensional data in Classification

Noppamas Akarachantachote
Huachiew Chalermprakiet University, Thailand

Abstract

This article reviews approaches of feature selection for high-dimensional data in classification task. Feature selection has become an interesting issue and needed in many areas of application. The progress of technology causes an ability of producing and collecting large datasets, especially high-dimension data. But only some of features influence predicting class. Therefore, it is important to select such features. This article summarizes approaches of feature selection including comparison among of them.

Keywords: classification, feature selection, high-dimensional data

ความนำ

การจำแนกประเภท (Classification) เป็นการวิเคราะห์ข้อมูลเพื่อการทำนายการเป็นสมาชิกกลุ่มโดยอาศัยข้อมูลตัวอย่างซึ่งทราบกลุ่ม และคุณลักษณะ (Feature) บางอย่างที่มีประโยชน์ต่อการทำนาย บางครั้งเรียกว่าเป็นการจำแนกประเภทแบบมีการสอน (Supervised classification) การกำหนดคุณลักษณะที่มีประโยชน์ได้อย่างเหมาะสมสำหรับการจำแนกประเภทจะทำให้การทำนายเกิดความแม่นยำ ข้อมูลที่มีคุณลักษณะเป็นจำนวนมากหรือเรียกว่าข้อมูลมีจำนวนมิติมาก (High-dimensional data) หากนำคุณลักษณะทั้งหมดไปวิเคราะห์การจำแนกประเภทอาจส่งผลเสียต่อการทำนายการเป็นสมาชิก กลุ่ม ข้อมูลในลักษณะนี้ เช่น ในการจำแนกผู้ป่วยที่เป็นมะเร็งกับไม่เป็น โดยอาศัยข้อมูลไมโครอาร์เรย์ (Microarray data) ซึ่งเป็นข้อมูลที่ได้จากการศึกษารูปแบบการแสดงออกของยีนของสิ่งมีชีวิตหลายยีน พร้อมๆ กัน โดยยีนที่ศึกษานี้มีจำนวนเป็นหลักพันหรือหมื่น แต่จำนวนยีนที่มีประโยชน์อาจมีเพียง 5% ของยีนทั้งหมดที่ศึกษา (Krzanowski & Hand, 2009) ข้อมูลสำหรับการจำแนกประเภทเอกสาร (Text classification) ก็นับเป็นข้อมูลที่มีจำนวนมิติมาก โดยคุณลักษณะที่อาจเป็นไปได้ในการใช้จำแนกเอกสารก็คือคำหรือวลีในเอกสาร ซึ่งมีจำนวนมาก ในแต่ละเอกสาร (Forman, 2003) และหากนำคุณลักษณะทั้งหมดที่มีจำนวนมากไปใช้ในการจำแนกประเภทเอกสารก็อาจส่งผลในแง่ลบต่อความแม่นยำในการทำนายได้ นอกจากนี้ข้อมูลทางดาราศาสตร์ที่ได้จากกล้องโทรทรรศน์ที่ใช้เทคโนโลยีขั้นสูงทำให้ได้ข้อมูลจากภาพของวัตถุซึ่งได้รับการวัดค่าพารามิเตอร์ในจำนวนหลักสิบหรือร้อย (Zheng & Zhang, 2008) การที่ข้อมูลเหล่านี้มีมิติเป็นจำนวนมากอาจก่อให้เกิดปัญหาหากนำไปใช้ค้นหารูปแบบโดยปราศจากการจัดการข้อมูลโดยอาศัยความรู้หรือข้อมูลเบื้องต้น เนื่องจากข้อมูลอาจเกิดการกระจายจนทำให้ในบางจุดอาจไม่มีข้อมูลอยู่เลย หากจำนวนตัวอย่างในการวิเคราะห์มีน้อยจนไม่เพียงพอต่อคุณลักษณะที่มีจำนวนมากอาจทำให้ได้ค่าประมาณที่ไม่ดี รวมทั้งทำให้เกิดการสิ้นเปลืองทรัพยากร เช่น เวลา หรือหน่วยความจำ สิ่งต่างๆ เหล่านี้เรียกว่าเป็นปัญหาของมิติข้อมูล

กระบวนการในการกรองคุณลักษณะที่มีจำนวนมากเหล่านี้จึงเป็นกระบวนการที่จำเป็นก่อนที่จะใช้วิธีการวิเคราะห์เพื่อจำแนกประเภท เพื่อช่วยลดมิติของข้อมูล กำจัดข้อมูลที่ไม่เกี่ยวข้อง และข้อมูลซ้ำซ้อน ทำให้เพิ่มความแม่นยำในการเรียนรู้ เพิ่มความเข้าใจต่อตัวแบบที่ได้ และสามารถลดเวลาในการเรียนรู้ข้อมูล อีกทั้งลดความต้องการของหน่วยความจำ กระบวนการนี้เรียกว่า การคัดเลือกคุณลักษณะ (Feature selection)

การคัดเลือกคุณลักษณะ (Feature selection)

การคัดเลือกคุณลักษณะได้รับการนิยามจากผู้เขียนหลายคน Dash and Liu (1997) ได้สรุปความหมายที่ครอบคลุมจากผู้เขียนหลายท่านไว้ว่า การคัดเลือกคุณลักษณะเป็นการเลือกเซตย่อยของคุณลักษณะที่มีขนาดเล็กที่สุด (ดีที่สุด) โดยสอดคล้องกับเงื่อนไขต่อไปนี้

1. ความแม่นยำของการจำแนกประเภทจะไม่ลดลงอย่างมีนัยสำคัญ
2. การกระจายของกลุ่ม (Class distribution) เมื่อใช้เฉพาะคุณลักษณะที่ถูกเลือกมีลักษณะใกล้เคียงกับการกระจายของกลุ่มเริ่มต้นเมื่อมีคุณลักษณะครบ

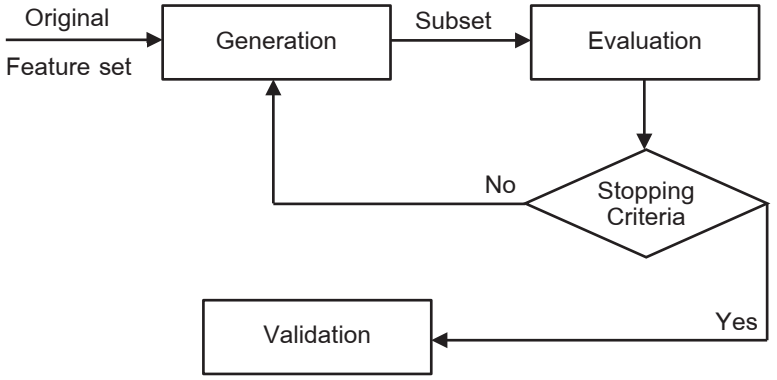
Kohavi and John (1996) ได้นิยามความหมายของคำว่า เซตย่อยของคุณลักษณะที่ดีที่สุด (Optimal feature subset) ว่าเป็นเซตย่อยของคุณลักษณะซึ่งมีความแม่นยำของการจำแนกประเภทสูงที่สุดในบรรดาเซตย่อยทั้งหมด

ไม่ว่าความหมายของการคัดเลือกคุณลักษณะจากผู้เขียนแต่ละคนจะเป็นอย่างไรก็ตาม แต่วัตถุประสงค์ของการคัดเลือกคุณลักษณะจะเป็นไปเพื่อ

1. ปรับปรุงประสิทธิภาพในการทำนาย
2. จัดเตรียมคุณลักษณะสำหรับการทำนายที่สามารถประมวลผลได้อย่างรวดเร็ว และมีประสิทธิภาพ
3. เพิ่มความเข้าใจต่อตัวแบบที่ได้

1. กระบวนการคัดเลือกคุณลักษณะโดยทั่วไป

กระบวนการคัดเลือกคุณลักษณะโดยทั่วไปประกอบด้วย 4 ขั้นตอน (ภาพที่ 1) ได้แก่ การสร้างเซตย่อย (Subset generation) การประเมินผล (Evaluation) เกณฑ์การหยุด (Stopping criterion) และการตรวจสอบ (Validation) โดยขั้นตอนวิธีการคัดเลือกคุณลักษณะเริ่มด้วยการสร้างเซตย่อยจากคุณลักษณะเริ่มต้นทั้งหมด จากนั้นประเมินผลเซตย่อยนั้น วนรอบซ้ำจนกระทั่งเป็นไปตามเกณฑ์การหยุดที่กำหนด แล้วจึงนำเซตย่อยที่ได้มาตรวจสอบโดยอาศัยขั้นตอนวิธีตัวจำแนกประเภท (Classifier algorithm)



ภาพที่ 1 กระบวนการคัดเลือกคุณลักษณะโดยทั่วไป (Novaković, Strbac, & Bulatović, 2011)

การสร้างเซตย่อย

การสร้างเซตย่อยเป็นกระบวนการสร้างย่อยของคุณลักษณะที่เป็นคู่แข่งโดยอาศัยฟังก์ชัน การประเมินค่าในการเลือกเซตย่อยใดๆ ซึ่งเซตย่อยทั้งหมดที่เป็นไปได้มีจำนวน 2^N เมื่อ N เป็นจำนวนคุณลักษณะของข้อมูลเริ่มต้น หาก N มีค่ามาก การใช้เทคนิคการค้นหาเซตย่อยทั้งหมด (Exhaustive search) จะไม่สามารถเป็นไปได้ เนื่องจากใช้เวลามากจนไม่อาจทำได้ในการปฏิบัติจริง ดังนั้นเทคนิค การค้นหาแบบฮิวริสติก (Heuristic search) ต่างๆ ถูกนำมาใช้เพื่อค้นหาเซตย่อย ซึ่งถึงแม้ว่าจะไม่ได้รับประกันผลที่ได้ว่าเป็นผลลัพธ์ที่ดีที่สุด แต่นับว่าได้ผลลัพธ์ที่ดีและสามารถทำได้จริงในการปฏิบัติ การเริ่มต้นการสร้างเซตย่อยสามารถทำได้โดย

1. การคัดเลือกแบบไปข้างหน้า (Forward selection) เป็นการสร้างเซตย่อยโดยเริ่มจากเซตว่าง แล้วเพิ่มคุณลักษณะที่ค้นหาได้และพบว่าผลการประเมินค่าสูงที่สุดเข้าสู่เซตย่อย

2 การกำจัดแบบย้อนกลับ (Backward elimination) เป็นการสร้างเซตย่อยโดยเริ่มจากคุณลักษณะทั้งหมด แล้วกำจัดคุณลักษณะที่ค้นหาได้และพบว่าผลการประเมินค่าน้อยที่สุด

3. การค้นหาแบบสุ่ม (Random search) เป็นการสร้างเซตย่อยโดยคุณลักษณะที่ถูกเพิ่มหรือกำจัดจะเป็นไปด้วยการสุ่ม

การประเมินผล

เซตย่อยแต่ละเซตที่ได้จากกระบวนการสร้างเซตย่อยต้องได้รับการประเมินผลโดยฟังก์ชันการประเมินค่า (Evaluation function) และเปรียบเทียบกับเซตย่อยที่ดีที่สุดก่อนหน้านี้โดยฟังก์ชันการประเมินค่าดังกล่าว ถ้าพบว่าเซตย่อยใหม่ดีกว่าเซตย่อยที่ดีที่สุดก่อนหน้านี้จะแทนที่ด้วยเซตย่อยใหม่

ฟังก์ชันการประเมินค่าเป็นเครื่องมือวัดความสามารถของคุณลักษณะหรือเซตย่อยของคุณลักษณะในการจำแนกกลุ่มที่แตกต่างกัน มีการแบ่งประเภทของฟังก์ชันการประเมินค่าในหลายลักษณะ โดย Dash and Liu (1997) ได้แบ่งออกเป็น 5 ประเภท ดังนี้

1. มาตรวัดระยะทาง (Distance measures)

มาตรวัดระยะทางจัดเป็นมาตรวัดการจำแนกประเภท (Discrimination measure) มาตรวัดความแตกต่าง (Divergence measure) หรือมาตรวัดการแยกออกจากกันได้ (Separability measure) ในปัญหาการจำแนกประเภทชนิดที่แบ่งเป็นสองกลุ่ม คุณลักษณะ X มีแนวโน้มได้รับเลือกมากกว่าคุณลักษณะ Y ถ้า X สามารถสรุปความแตกต่างระหว่างสองกลุ่มได้มากกว่า Y

2. มาตรวัดสารสนเทศ (Information measures)

มาตรวัดชนิดนี้จะกำหนดการได้รับสารสนเทศ (Information gain) จากคุณลักษณะการได้รับสารสนเทศจากคุณลักษณะ X หมายถึงความแตกต่างระหว่างความไม่แน่นอนก่อนได้รับสารสนเทศจาก X (Prior uncertainty) และค่าคาดหวังของความไม่แน่นอนหลังได้รับสารสนเทศจาก X (Expected posterior uncertainty) คุณลักษณะ X มีโอกาสถูกเลือกมากกว่า Y ถ้าสารสนเทศที่ได้รับจาก X มีมากกว่าที่ได้รับจาก Y

3. มาตรวัดความไม่เป็นอิสระ (Dependence measures)

มาตรวัดความไม่เป็นอิสระ หรือมาตรวัดความสัมพันธ์ (Correlation measure) ใช้วัดความสามารถในการทำนายค่าของตัวแปรหนึ่งจากอีกตัวแปรหนึ่ง ถ้าความสัมพันธ์ของคุณลักษณะ X กับกลุ่ม C สูงกว่าความสัมพันธ์ของคุณลักษณะ Y กับกลุ่ม C แล้ว จะได้ว่าคุณลักษณะ X มีโอกาสได้รับเลือกมากกว่าคุณลักษณะ Y นอกจากการใช้เป็นมาตรวัดเพื่อคัดเลือกคุณลักษณะที่มีความสัมพันธ์กับตัวแปรกลุ่มแล้ว มาตรวัดความไม่เป็นอิสระยังสามารถนำมาวัดระดับความซ้ำซ้อนของคุณลักษณะได้ด้วย

4. มาตรวัดความคงเส้นคงวา (Consistency measures)

มาตรวัดชนิดนี้จะทำการหาเซตย่อยที่เล็กที่สุดที่อัตราความไม่คงเส้นคงวา (Inconsistency rate) มีค่ายอมรับได้ ความไม่คงเส้นคงวาหมายถึงการที่สองตัวอย่างใดๆ มีค่าของคุณลักษณะที่เหมือนกันแต่อยู่ต่างกลุ่มกัน

5. มาตรวัดอัตราความผิดพลาดของตัวจำแนกประเภท (Classifier error rate measures)

มาตรวัดชนิดนี้อาศัยการใช้ขั้นตอนวิธีตัวจำแนกประเภท (Classifier algorithm) หรือขั้นตอนวิธีการเรียนรู้ (Learning algorithm) เพื่อจำแนกประเภทในการวัดความผิดพลาด หรือความแม่นยำเมื่อใช้เซตย่อยในการจำแนก เป็นเสมือนฟังก์ชันในการประเมินค่าของเซตย่อยนั้น

สิ่งที่ควรพิจารณาสำหรับการเลือกฟังก์ชันการประเมินค่าได้แก่ ความสามารถในการใช้กับตัวจำแนกประเภทโดยทั่วไป (Generality) ความซับซ้อนของเวลาที่ใช้ (Time complexity) และความแม่นยำ (Accuracy) โดย Dash and Liu (1997) ได้สรุปเปรียบเทียบความสามารถของฟังก์ชันการประเมินค่าในแต่ละมาตรวัดดังนี้

ตารางที่ 1 การเปรียบเทียบฟังก์ชันการประเมินค่า

ฟังก์ชันการประเมินค่า	Generality	Time complexity	Accuracy*
มาตรวัดระยะทาง	✓	ต่ำ	-
มาตรวัดสารสนเทศ	✓	ต่ำ	-
มาตรวัดความไม่เป็นอิสระ	✓	ต่ำ	-
มาตรวัดความคงเส้นคงวา	✓	ปานกลาง	-
มาตรวัดอัตราความผิดพลาดของตัวจำแนกประเภท	✗	สูง	สูงมาก

หมายเหตุ. * ไม่สามารถสรุปความแม่นยำของฟังก์ชันการประเมินค่าได้ ยกเว้นมาตรวัดอัตราความผิดพลาดของตัวจำแนกประเภท เนื่องจากขึ้นอยู่กับชุดข้อมูล และตัวจำแนกประเภทที่ใช้หลังจากการคัดเลือกคุณลักษณะ

เกณฑ์การหยุด

การกำหนดเกณฑ์ที่ใช้ในการหยุดค้นหาเซตย่อยขึ้นอยู่กับกระบวนการสร้างเซตย่อย ซึ่งอาจหยุดเมื่อ

1. ครบกำหนดตามจำนวนคุณลักษณะที่ต้องการ
2. ครบกำหนดจำนวนรอบของการทำซ้ำที่ต้องการ

นอกจากนี้ เกณฑ์การหยุดยังขึ้นอยู่กับฟังก์ชันการประเมินค่า ซึ่งอาจกำหนดให้หยุดเมื่อ

1. การเพิ่มขึ้นหรือการลดลงของคุณลักษณะใดๆ ไม่ทำให้ได้เซตย่อยที่ดีกว่า
2. เซตย่อยที่ได้มีความเหมาะสมสอดคล้องตามเกณฑ์ของฟังก์ชันการประเมินค่า

การตรวจสอบ

ขั้นตอนการตรวจสอบจริงๆ แล้วไม่ได้เป็นส่วนหนึ่งของการคัดเลือกคุณลักษณะ แต่ในทางปฏิบัติจะต้องมีการตรวจสอบผลที่ได้จากวิธีการคัดเลือกคุณลักษณะนั้นๆ ซึ่งเป็นขั้นตอนของการทดสอบความถูกต้องเหมาะสมของเซตย่อยที่ได้โดยทดสอบหลายครั้งและเปรียบเทียบผลกับวิธีคัดเลือกคุณลักษณะอื่นๆ หรือเปรียบเทียบกับเมื่อไม่ใช้วิธีการคัดเลือกคุณลักษณะ โดยอาศัยชุดข้อมูลเทียม (Artificial datasets) หรือชุดข้อมูลจริง (Real-world datasets)

2. แนวทางของการคัดเลือกคุณลักษณะ

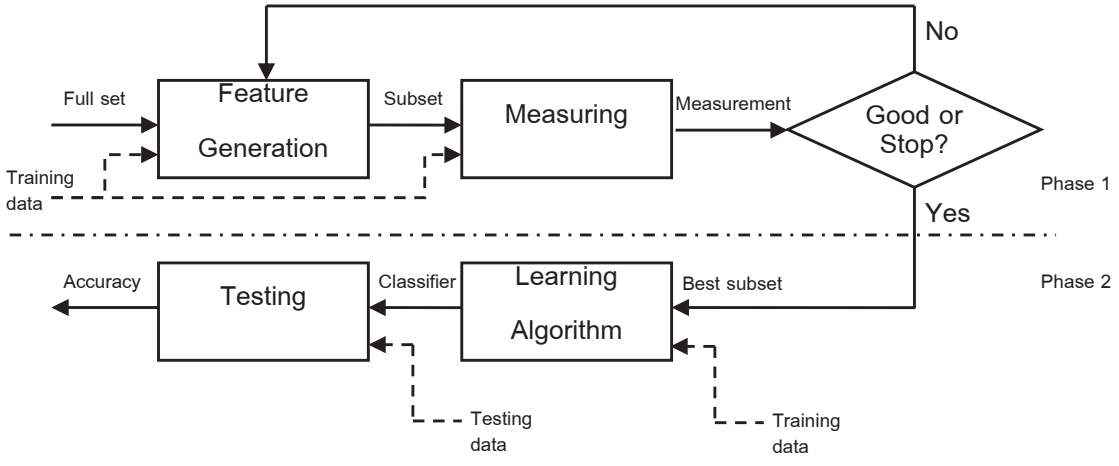
ในบริบทของการจำแนกประเภท การคัดเลือกคุณลักษณะสามารถแบ่งออกได้เป็น 3 วิธี ได้แก่ วิธีฝังตัว (Embedded methods) วิธีฟิลเตอร์ (Filter methods) และวิธีแรปเปอร์ (Wrapper methods)

วิธีฝังตัว เป็นวิธีที่การคัดเลือกคุณลักษณะเป็นส่วนหนึ่งอยู่ในกระบวนการจำแนกประเภทด้วย เช่น กระบวนการสร้างต้นไม้การตัดสินใจ (Decision trees) จะทำการเลือกเซตย่อยที่มีคุณลักษณะที่เหมาะสมในแต่ละ

ขั้นตอนวิธี เนื่องจากวิธีแบบฝังตัวนี้จะมีขั้นตอนวิธีที่เฉพาะเจาะจงในแต่ละขั้นตอนการเรียนรู้เพื่อจำแนกประเภท ดังนั้นจึงขอนำเสนอรายละเอียดเฉพาะวิธีฟิลเตอร์ และวิธีแรปเปอร์ เท่านั้น

2.1 วิธีฟิลเตอร์

วิธีฟิลเตอร์เป็นวิธีการคัดเลือกคุณลักษณะที่เร็วและง่ายต่อการตีความ โดยจะกำจัดคุณลักษณะที่ไม่เกี่ยวข้อง (Irrelevant feature) ต่อการจำแนกประเภทด้วยคุณสมบัติในเนื้อแท้ของข้อมูล ซึ่งวัดเป็นคะแนน และเรียงลำดับคุณลักษณะตามคะแนนที่ได้ โดยส่วนใหญ่ถ้าคุณลักษณะใดที่มีคะแนนต่ำจะถูกกำจัด คุณสมบัตินี้ก็คือฟังก์ชันการประเมินค่าที่กล่าวในตอนต้นนั่นเอง สำหรับวิธีฟิลเตอร์นั้นสามารถใช้ฟังก์ชันการประเมินค่าด้วยมาตรวัดระยะทาง มาตรวัดสารสนเทศ มาตรวัดความไม่เป็นอิสระ หรือมาตรวัดความคงเส้นคงวา การคัดเลือกคุณลักษณะโดยวิธีฟิลเตอร์มีกระบวนการที่เป็นอิสระจากขั้นตอนวิธีการเรียนรู้เพื่อจำแนกประเภท ซึ่งแสดงดังภาพที่ 2



ภาพที่ 2 การคัดเลือกคุณลักษณะโดยวิธีฟิลเตอร์ (Liu & Motoda, 1998)

การคัดเลือกคุณลักษณะโดยวิธีฟิลเตอร์ประกอบด้วยสองช่วง ช่วงที่ 1 เป็นการวัดค่าความสามารถของคุณลักษณะโดยอาศัยฟังก์ชันการประเมินค่า โดยไม่มีการจำแนกประเภทในช่วงนี้ ส่วนช่วงที่ 2 เป็นการเรียนรู้เพื่อสร้างตัวจำแนกประเภทบนข้อมูลฝึกฝน (Training data) ด้วยคุณลักษณะที่ถูกเลือกมาจากช่วงที่ 1 จากนั้นทดสอบตัวจำแนกประเภทที่ได้โดยพิจารณาจากความแม่นยำบนข้อมูลทดสอบ (Test data)

ฟังก์ชันการประเมินค่าถูกนำมาใช้วัดความสามารถของคุณลักษณะหรือเซตย่อยของคุณลักษณะ ซึ่งฟังก์ชันที่ใช้สำหรับวิธีฟิลเตอร์มีดังนี้

มาตรวัดระยะทาง

ฟังก์ชันการประเมินค่าที่อยู่ในมาตรวัดระยะทางเช่น ค่าวัดระยะทางยูคลิเดียน (Euclidean distance measure) และค่าวัดระยะทางมหาลาโนบิส (Mahalanobis distance measure) เป็นต้น

ค่าวัดระยะทางระหว่างข้อมูลตัวอย่างที่ p และ q สำหรับ n คุณลักษณะ (มิติ) โดยที่ Pk เป็นคุณลักษณะที่ k ของตัวอย่างที่ p และ q ตามลำดับ ถ้าวัดด้วยค่าวัดระยะทางยูคลิเดียนเป็นดังนี้

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_k - q_k)^2} \quad \ldots (1)$$

ถ้าวัดด้วยค่าวัดระยะทางมหาลาโนบิสเป็นดังนี้

$$d(p,q) = \sqrt{((p-q)^{\dagger}T\ (^{\dagger}(-1)\ (p-q))} \quad \ldots (2)$$

ขั้นตอนวิธีการคัดเลือกคุณลักษณะที่นำมาตรวัดระยะทางสำหรับการเป็นฟังก์ชันการประเมินค่า ที่มีความโดดเด่นคือขั้นตอนวิธีรีลีฟ (Relief algorithm) (Dash & Liu, 1997) ซึ่งนำเสนอโดย Kira and Rendell (1992) รีลีฟ เป็นขั้นตอนวิธีในการหาค่าเพื่อประเมินผลคุณลักษณะโดยเริ่มจากการเลือกตัวอย่างอย่างสุ่มตามจำนวนที่ผู้ใช้กำหนด สำหรับแต่ละตัวอย่างจะทำการหาตัวอย่างที่อยู่ใกล้ตัวอย่างนั้นที่สุดและอยู่กลุ่มเดียวกัน (Nearest hit) และหาตัวอย่างที่อยู่ใกล้ที่สุดแต่อยู่คนละกลุ่ม (Nearest miss) โดยวัดค่าความใกล้ด้วยค่าวัดระยะทางยูคลิเดียน (Euclidean distance measure) คุณลักษณะใดที่สามารถทำให้เห็นความแตกต่างระหว่างตัวอย่างใดๆ กับตัวอย่างที่อยู่ใกล้ที่สุดแต่อยู่คนละกลุ่มได้มากกว่า แสดงว่าคุณลักษณะนั้นจะเป็นคุณลักษณะที่มีความเกี่ยวข้อง (Relevant feature) ในการจำแนกประเภทได้มากกว่า ส่วนคุณลักษณะใดที่ทำให้เห็นความแตกต่างระหว่างตัวอย่างใดๆ กับตัวอย่างที่อยู่ใกล้ที่สุดและอยู่กลุ่มเดียวกันได้มากกว่าแสดงว่าคุณลักษณะนั้นจะเป็นคุณลักษณะที่มีความเกี่ยวข้องน้อยกว่า หลังจากที่วัดค่าครบทุกตัวอย่างตามที่กำหนดแล้วนำมาคำนวณเป็นค่าน้ำหนักของแต่ละคุณลักษณะ คุณลักษณะที่มีค่าน้ำหนักเกินที่กำหนดจะถูกเลือก แนวทางหนึ่งสำหรับกำหนดเกณฑ์ทำได้โดยเลือกคุณลักษณะที่มีค่าน้ำหนักเป็นบวก (Dash & Liu, 1997)

ขั้นตอนวิธีรีลีฟมีข้อจำกัดสำหรับจำนวนกลุ่มซึ่งใช้ได้เฉพาะกรณีสองกลุ่ม (Binary Classes) ข้อจำกัดดังกล่าวสามารถแก้ไขได้โดยใช้ขั้นตอนวิธีรีลีฟเอฟ (ReliefF algorithm) ซึ่งพัฒนาโดย Kononenko (1994) ขั้นตอนวิธีรีลีฟเอฟนอกจากสามารถใช้คัดเลือกคุณลักษณะเพื่อจำแนกประเภทข้อมูลที่มีมากกว่าสองกลุ่มได้แล้ว ยังช่วยแก้ปัญหาในเรื่องความไม่สมบูรณ์ของข้อมูลได้อีกด้วย

มาตรวัดสารสนเทศ

มาตรวัดสารสนเทศสนใจความสามารถในการแบ่งแยกกลุ่มโดยพิจารณาจากการได้รับสารสนเทศ (Information gain) ซึ่งอาศัยค่าเอนโทรปี (Entropy) ในการวัด เอนโทรปีเป็นค่าวัดสารสนเทศเชิงทฤษฎี (Information-theoretic measure) ของความไม่แน่นอนบนชุดข้อมูลฝึกฝนอันเนื่องจากการมีกลุ่มที่เป็นไปได้ที่มากกว่าหนึ่งกลุ่ม (Bramer, 2007) หรือเป็นค่าที่ใช้วัดระดับการสุ่มหรือความไม่เป็นระเบียบของเหตุการณ์ซึ่งคำนวณโดย

กำหนดให้ Y เป็นตัวแปรแทนกลุ่มซึ่งมี k กลุ่ม และ *X* เป็นตัวแปรแทนคุณลักษณะ ซึ่งคุณลักษณะ *X* แบ่งออกเป็น *l* ประเภท เอนโทรปีของ *Y* เขียนแทนด้วย *H(Y)* หาได้จาก

$$H(Y) = -\sum_{i=1}^{k} P(Y = y_i)\log_2(P(Y = y_i)) \quad \ldots (3)$$

เอนโทรปีอย่างมีเงื่อนไขของ Y โดยกำหนด X เขียนแทนด้วย H(Y|X) หาได้จาก

$$H(Y|X) = -\sum_{j=1}^{l} P(X = x_i)H(Y|X = x_i) \quad \ldots\ldots (4)$$

การลดลงของเอนโทรปี หรือเรียกว่าเป็นสารสนเทศที่ได้รับจากแต่ละคุณลักษณะ เขียนแทนด้วย *IG(Y;X)* ซึ่งหาได้จาก

$$IG(Y;X) = H(Y) - H(Y|X) \quad \ldots\ldots (5)$$

สารสนเทศที่ได้รับจากคุณลักษณะใดมีค่าสูงกว่า แสดงว่าคุณลักษณะนั้นสามารถแบ่งแยกความแตกต่างระหว่างกลุ่มได้ดีกว่า

สารสนเทศที่ได้รับมักมีค่ามากถ้าคุณลักษณะใดๆ แบ่งออกเป็นหลายประเภท ทั้งที่อาจไม่ใช่คุณลักษณะที่เกี่ยวข้องกับการจำแนกประเภท เพื่อแก้ไขปัญหาดังกล่าว Ross Quinlan (1993, อ้างถึงใน Bramer, 2007) จึงมีการปรับค่าดังกล่าวด้วยสารสนเทศของการแบ่งแยก (Split information) ได้ค่าเป็นอัตราส่วนเกน (Gain ratio) ซึ่งเขาได้นำไปใช้กับตัวจำแนกประเภท C4.5 อัตราส่วนเกนคำนวณได้ดังนี้

$$GR(Y;X) = \frac{IG(Y;X)}{splitinfo(X)} \quad \ldots\ldots (6)$$

โดยที่ *GR(Y;X)* แทนอัตราส่วนเกนของคุณลักษณะ *X* ต่อการจำแนกประเภทตัวแปร *Y*

Splitinfo *(X)* แทนสารสนเทศของการแบ่งแยกของคุณลักษณะ *X* ซึ่งคำนวณได้จาก

$$splitinfo(X) = -\sum_{i=1}^{l} P(X = x_i)\log_2 P(X = x_i) \quad \ldots\ldots (7)$$

มาตรวัดความไม่เป็นอิสระ

ค่าสถิติไค-สแควร์ (Chi-square statistic) เป็นค่าสถิติที่ใช้ทดสอบความสัมพันธ์ของตัวแปรเชิงคุณภาพสองตัวแปร ซึ่งสามารถนำมาใช้ในการคัดเลือกคุณลักษณะได้ โดยขนาดความสัมพันธ์ระหว่างคุณลักษณะใดๆ กับตัวแปรกลุ่มสามารถวัดด้วยค่าสถิติไค-แสควร์ ซึ่งคำนวณดังนี้

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{\left(M_{ij} - m_{ij}\right)^2}{m_{ij}} \quad \ldots\ldots (8)$$

โดยที่ m_{ij} เป็นความถี่คาดหวังเมื่อ *X* และ *Y* เป็นอิสระกัน ซึ่ง $m_{ij} = \frac{M_{i.}M_{.j}}{N}$ ถ้า *X* และ *Y* เป็นอิสระกันอย่างสมบูรณ์จะได้ว่า $M_{ij} = m_{ij}$ ดังนั้นถ้าทั้งสองค่านี้มีความแตกต่างกันมากจะแสดงถึงความสัมพันธ์ที่มากของทั้งสองตัวแปร หากกำหนดให้ *X* เป็นตัวแปรแทนคุณลักษณะ *Y* เป็นตัวแปรกลุ่ม คุณลักษณะใดที่มีค่า ไค-สแควร์มากกว่าจะแสดงถึงความเกี่ยวข้องของคุณลักษณะนั้นที่มีต่อตัวแปรกลุ่ม *Y*

ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson correlation coefficient) เป็นค่าที่ใช้วัดระดับความสัมพันธ์เชิงเส้นระหว่างสองตัวแปร มีค่าอยู่ตั้งแต่ -1 ถึง +1 ซึ่งแสดงถึงความสัมพันธ์กันอย่างสมบูรณ์ทางลบและทางบวกตามลำดับ และค่าเท่ากับ 0 แสดงถึงการไม่สัมพันธ์กัน แสดงการคำนวณดังนี้

$$R(i) = \frac{cov(X_i, Y)}{\sqrt{var(X_i)var(Y)}} \quad \ldots\ldots (9)$$

โดยที่ $R(i)$ แทนสัมประสิทธิ์สหสัมพันธ์เพียร์สันของคุณลักษณะที่ i

$cov(X_i, Y)$ แทนความแปรปรวนร่วมของคุณลักษณะที่ i และตัวแปรกลุ่ม Y

$var(X_i), var(Y)$ แทนความแปรปรวนของคุณลักษณะที่ i และตัวแปรกลุ่ม Y ตามลำดับ

Hall (1999) ได้พัฒนาขั้นตอนวิธีการคัดเลือกคุณลักษณะบนพื้นฐานของความสัมพันธ์ (Correlation-based feature selection: CFS) ในการประเมินค่าเซตย่อยของคุณลักษณะ โดยมีแนวความคิดว่าเซตย่อยของคุณลักษณะที่ดีจะต้องประกอบด้วยคุณลักษณะที่มีความสัมพันธ์อย่างสูงกับตัวแปรกลุ่ม และไม่มีความสัมพันธ์กันเอง ซึ่ง Hall ได้กำหนดค่าที่ใช้วัดความสามารถของเซตย่อยขนาด k โดยอาศัยค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันในการคำนวณค่าดังนี้

$$r_{x_k y} = \frac{k\overline{r_{ky}}}{\sqrt{k + k(k-1)\overline{r_{kk}}}} \quad \ldots\ldots\ldots\ldots(10)$$

โดยที่

$\overline{r_{ky}}$ แทนค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ระหว่างแต่ละคุณลักษณะกับตัวแปรกลุ่ม

$\overline{r_{kk}}$ แทนค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ภายในระหว่างแต่ละคุณลักษณะ

มาตรวัดความคงเส้นคงวา

มาตรวัดความคงเส้นคงวาวัดค่าด้วยอัตราความไม่คงเส้นคงวา (Inconsistency rate) ซึ่งวัดความไม่คงเส้นคงวาสำหรับเซตย่อยของคุณลักษณะใดๆ แนวทางการคำนวณเป็นดังนี้ (Dash & Liu, 2003)

ตัวอย่างใดๆ อย่างน้อยสองตัวอย่างที่มีรูปแบบ (Pattern) หรือค่าของคุณลักษณะในเซตย่อยนั้นที่เหมือนกันแต่อยู่ต่างกลุ่มกัน จะถูกเรียกว่าตัวอย่างนั้นๆ มีรูปแบบที่ไม่คงเส้นคงวา จำนวนความไม่คงเส้นคงวา (Inconsistency count) ของรูปแบบสำหรับเซตย่อยใดๆ หาได้จาก จำนวนตัวอย่างที่เป็นไปตามรูปแบบนั้นลบด้วยจำนวนที่มากที่สุดของกลุ่มต่างๆ เช่น กำหนดให้เซตย่อย S ของคุณลักษณะ ประกอบด้วย p รูปแบบ ซึ่งแต่ละรูปแบบมี n_p ตัวอย่าง โดยจำนวนตัวอย่างในรูปแบบนั้นอยู่กลุ่มที่ 1 2 และ 3 จำนวน c_1 c_2 และ c_3 ตามลำดับ *(ซึ่ง* $c_1+c_2+c_3= n_p$*)* สมมติว่าในรูปแบบที่ p นั้น c_3 มีจำนวนสูงที่สุด จะได้ว่าจำนวนความไม่คงเส้นคงวาสำหรับรูปแบบที่ p จะเท่ากับ $n_p - c_3$ อัตราความไม่คงเส้นคงวาของเซตย่อย S หาได้จาก

$$IR(S) = \frac{\sum_{i=1}^{p} IC_i}{N} \quad \ldots\ldots\ldots\ldots(11)$$

โดยที่

IC_i แทนจำนวนความไม่คงเส้นคงวาของรูปแบบที่ i

N แทนจำนวนตัวอย่างทั้งหมด

ถ้าอัตราความไม่คงเส้นคงวาสำหรับเซตย่อยของคุณลักษณะใดๆ มีค่าน้อยกว่าหรือเท่ากับเกณฑ์ที่กำหนด จะเรียกว่าเซตย่อยนั้นมีความคงเส้นคงวา (Consistency)

Liu and Setiono ได้ใช้ขั้นตอนวิธีแอลวีเอฟ (LVF algorithm) ซึ่งใช้มาตรวัดความคงเส้นคงวาในการประเมินค่า โดยเริ่มจากการสร้างเซตย่อย S ของคุณลักษณะอย่างสุ่ม ถ้า S ประกอบไปด้วยคุณลักษณะที่มีจำนวนน้อยกว่าเซตย่อยที่ดีที่สุดในปัจจุบันแล้วจะทำการเปรียบเทียบอัตราความไม่คงเส้นคงวาของ S กับเซตย่อยที่ดีที่สุดในขณะนั้น ถ้า S มีความคงเส้นคงวาอย่างน้อยเท่ากับของเซตย่อยที่ดีที่สุด จะทำการแทนที่เซตย่อยที่ดีที่สุดด้วยเซตย่อย S (Liu & Setiono อ้างถึงใน Hall, 1999)

ถ้าพิจารณาจากกลุ่มของคุณลักษณะที่ประเมิน วิธีการคัดเลือกคุณลักษณะโดยวิธีฟิลเตอร์สามารถแบ่งประเภทได้เป็น 2 ประเภท

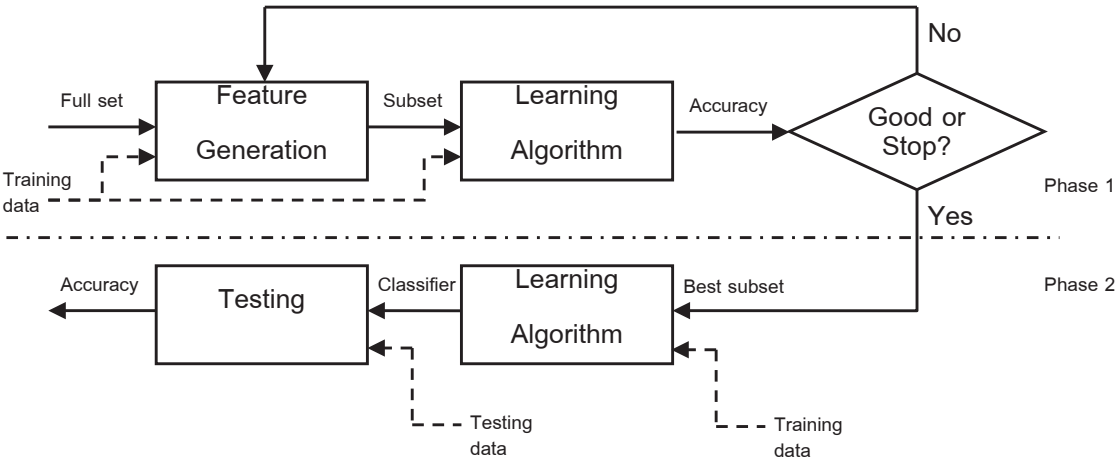
1. วิธีฟิลเตอร์แบบคุณลักษณะเดี่ยว (Univariate filter method) เป็นวิธีการคัดเลือกคุณลักษณะโดยประเมินผลทีละคุณลักษณะแยกจากกัน จากนั้นเรียงลำดับคุณลักษณะตามค่าที่ได้จากฟังก์ชัน การประเมินค่าเซตย่อยของคุณลักษณะที่ถูกเลือกจะได้จากการเลือกคุณลักษณะที่มีค่าสูงสุดจำนวน *N* คุณลักษณะ หรือเลือกคุณลักษณะที่มีค่าสูงกว่าเกณฑ์ *t* โดยที่ *N* และ *t* เป็นเกณฑ์ที่ผู้ใช้เป็นผู้กำหนดวิธีนี้เป็นวิธีที่ทำได้ง่ายและประมวลผลได้รวดเร็ว แต่มีข้อเสียในเรื่องการละเลยความสัมพันธ์ระหว่างคุณลักษณะซึ่งอาจนำไปสู่การลดประสิทธิภาพในการจำแนกประเภท

2. วิธีฟิลเตอร์แบบหลายคุณลักษณะ (Multivariate filter method) เป็นวิธีการคัดเลือกคุณลักษณะโดยประเมินผลเซตย่อยของคุณลักษณะ เพื่อเลือกเซตย่อยที่เหมาะสมที่สุด วิธีนี้ช่วยแก้ปัญหาที่เป็นข้อเสียของวิธีฟิลเตอร์แบบคุณลักษณะเดี่ยว โดยในการคัดเลือกได้พิจารณาความสัมพันธ์ระหว่างคุณลักษณะในระดับหนึ่งด้วย

2.2 วิธีแรปเปอร์

John, Kohavi, and Pfleger (1994) เป็นผู้แรกที่สนับสนุนการใช้แรปเปอร์เป็นการคัดเลือกคุณลักษณะ ซึ่งได้ให้คำนิยามของความเกี่ยวข้องของคุณลักษณะ (Feature relevance) และกล่าวว่า แรปเปอร์เป็นวิธีที่สามารถค้นพบคุณลักษณะที่มีความเกี่ยวข้อง (Relevant features)ได้ คุณลักษณะ X_i จะเรียกว่ามีความเกี่ยวข้องกับตัวแปรกลุ่ม หรือตัวแปรเป้าหมาย (Target) อย่างเข้ม (Strongly relevant) ก็ต่อเมื่อการแจกแจงความน่าจะเป็นอย่างมีเงื่อนไขของกลุ่มเมื่อกำหนดคุณลักษณะทั้งหมดนั้นเปลี่ยนไป ถ้า X_i ถูกกำจัด และ คุณลักษณะ X_i จะเรียกว่ามีความเกี่ยวข้องกับตัวแปรกลุ่มอย่างอ่อน (Weakly relevant) ถ้าการแจกแจงความน่าจะเป็นอย่างมีเงื่อนไขของกลุ่มเมื่อกำหนดเซตย่อยของคุณลักษณะ *S* (ซึ่งรวม X_i) นั้นเปลี่ยนไป ถ้า X_i ถูกกำจัด คุณลักษณะใดที่ไม่ได้มีความเกี่ยวข้องอย่างเข้มหรืออย่างอ่อน จะเรียกว่าเป็นคุณลักษณะที่ไม่มีความเกี่ยวข้อง (Irrelevant feature)

การคัดเลือกคุณลักษณะโดยวิธีแรปเปอร์อาศัยขั้นตอนวิธีการเรียนรู้ในการประเมินค่าเซตย่อย ซึ่งจะทำให้ได้เซตย่อยที่มีความแม่นยำในการจำแนกประเภทมากกว่าการใช้ค่าจากมาตรวัดอื่นในการประเมินค่า ขั้นตอนการทำงานประกอบด้วยสองช่วง (ภาพที่ 3) ช่วงที่ 1 เป็นการคัดเลือกเซตย่อยของคุณลักษณะซึ่งเลือกเซตย่อยที่ดีที่สุดโดยดูจากความแม่นยำของตัวจำแนกประเภท (บนข้อมูลฝึกฝน) หรือใช้มาตรวัดอัตราความผิดพลาดของตัวจำแนกประเภทเป็นฟังก์ชันการประเมินค่า ส่วนช่วงที่ 2 เป็นการเรียนรู้และการทดสอบ โดยนำเซตย่อยที่ดีที่สุดที่ได้จากการคัดเลือกคุณลักษณะในช่วงแรกมาเรียนรู้เพื่อสร้างตัวแบบบนข้อมูลฝึกฝน และทำการทดสอบตัวแบบที่ได้บนข้อมูลทดสอบ



ภาพที่ 3 การคัดเลือกคุณลักษณะโดยวิธีแรปเปอร์ (Liu & Motoda, 1998, pp.34)

เมื่อแต่ละเซตย่อยถูกสร้าง ตัวแบบจะถูกสร้างจากข้อมูลของเซตย่อยนั้น และคำนวณหาค่าอัตราความผิดพลาด โดยเซตย่อยที่มีความแม่นยำสูงสุด (อัตราความผิดพลาดต่ำสุด) จะถูกเก็บไว้ เมื่อกระบวนการคัดเลือกสิ้นสุด เซตย่อยที่มีความแม่นยำสูงสุดจะถูกเลือก ช่วงที่ 2 เป็นกระบวนการเรียนรู้ และการทดสอบ ตามขั้นตอนปกติของการจำแนกประเภท ซึ่งจะต้องหาค่าความแม่นยำในการทำนายบนข้อมูลทดสอบ

ความแม่นยำของตัวแบบบนข้อมูลฝึกฝนอาจไม่สะท้อนความแม่นยำบนข้อมูลทดสอบ วิธีหนึ่งที่จะแก้ปัญหานี้สามารถทำได้โดยใช้การตรวจสอบไขว้ (Cross validation) ซึ่งการทำเช่นนี้จะทำให้กระบวนการคัดเลือกคุณลักษณะใช้เวลามากขึ้น รวมทั้งกระบวนการเดิมของวิธีแรปเปอร์ซึ่งใช้เวลามาก อยู่แล้ว เนื่องจากมีการเรียนรู้หลายครั้งเท่ากับจำนวนครั้งที่สร้างเซตย่อยซึ่งจะทำให้ใช้เวลามาก ดังนั้น นักวิจัยมักใช้ขั้นตอนวิธีการเรียนรู้แบบฮิวริสติก เช่น ตัวจำแนกประเภทนาอีฟเบย์ (Naïve Bayes) หรือ ตัวเรียนรู้ต้นไม้การตัดสินใจ (Decision tree) (Kohavi & John, 1998, อ้างถึงใน Liu & Motoda, 1998) เป็นต้น การประยุกต์วิธีแรปเปอร์จะแตกต่างกันไปในเรื่องขั้นตอนวิธีการเรียนรู้ และเทคนิคการค้นหาที่ใช้

วิธีแรปเปอร์สำหรับตัวเรียนรู้ต้นไม้การตัดสินใจ (Wrappers for decision tree learner)

John, Kohavi, and Pfleger (1994) ได้ใช้ ID3 และ C4.5 ซึ่งเป็นตัวเรียนรู้ต้นไม้การตัดสินใจ (Decision tree learning algorithm) ในการทดลองของเขาได้ใช้ทั้งข้อมูลเทียม และข้อมูลจริง และใช้การเริ่มต้นสร้างเซตย่อยแบบไปข้างหน้า และแบบย้อนกลับ

Vafaie and De Jong (1995) ได้ใช้เทคนิคการค้นหาแบบจีเนติก (Genetic search strategies) ในการคัดเลือกคุณลักษณะโดยวิธีแรปเปอร์ สำหรับปรับปรุงประสิทธิภาพตัวเรียนรู้ต้นไม้การตัดสินใจ ในปัญหาการจำแนกประเภทเนื้อผ้า

วิธีแรปเปอร์สำหรับตัวจำแนกประเภทแบบเบย์ (Wrappers for Bayes classifiers)

Langley and Sage (1994) ได้ปรับปรุงตัวจำแนกประเภทนาอีฟเบย์ (Naïve bayes classifier) โดยกำจัดคุณลักษณะที่มีความซ้ำซ้อน (Redundant feature) เนื่องจากข้อสมมติเบื้องต้นของตัวจำแนกประเภทนาอีฟเบย์คือการแจกแจงความน่าจะเป็นของแต่ละคุณลักษณะจะต้องเป็นอิสระต่อกันในแต่ละกลุ่ม ดังนั้นจึงได้นำการสร้างเซตย่อยแบบไปข้างหน้ามาใช้สำหรับตัวจำแนกประเภทนาอีฟเบย์ เนื่องจากการทำเช่นนี้จะสามารถตรวจจับคุณลักษณะที่ซ้ำซ้อนที่จะถูกนำเข้าสู่เซตย่อย และใช้เทคนิคการค้นหาแบบละโมภ (Greedy search) ซึ่งเป็นการค้นหาแบบฮิวริสติก

Pazzani (1995) ได้นำการคัดเลือกคุณลักษณะมาปรับปรุงตัวจำแนกประเภทนาอีฟเบย์ โดยสร้างเซตย่อยแบบไปข้างหน้าและแบบย้อนกลับด้วยเทคนิคการค้นหาแบบปีนเขา (Hill climbing search strategy) ซึ่งเป็นการค้นหาแบบฮิวริสติก โดยมีการเพิ่มขั้นตอนสำหรับการสร้างเซตย่อยแบบไปข้างหน้า นอกจากจะเพิ่มคุณลักษณะทีละตัวสู่เซตย่อยแล้วยังมีการสร้างคุณลักษณะใหม่โดยรวมคุณลักษณะเป็นคู่จากคุณลักษณะที่ยังไม่ถูกเลือก และคุณลักษณะที่ถูกเลือกแล้ว ส่วนการสร้างเซตย่อยแบบย้อนกลับได้มีเพิ่มขั้นตอนการแทนที่คู่ของคุณลักษณะด้วยคุณลักษณะตัวหนึ่งในคู่นั้น นอกเหนือจากกระบวนการเดิมที่กำจัดคุณลักษณะทีละตัวออกจากเซตย่อย

การเปรียบเทียบวิธีการคัดเลือกคุณลักษณะและแนวทางการคัดเลือกคุณลักษณะในอนาคต

วิธีแรปเปอร์เป็นวิธีการคัดเลือกคุณลักษณะที่มีความซับซ้อนในการคำนวณสูงสุด ตามมาด้วยวิธีฝังตัว (Janecek, 2009) ทั้งสองวิธีดังกล่าวมีแนวทางในการคัดเลือกเซตย่อยของคุณลักษณะบนพื้นฐานของขั้นตอนวิธีการเรียนรู้ที่เฉพาะเจาะจงซึ่งมีแนวโน้มจะทำให้เกิด Overfitting มากกว่าวิธีฟิลเตอร์ซึ่งเป็นอิสระจากขั้นตอนวิธีการเรียนรู้ แต่วิธีแรปเปอร์ และวิธีฝังตัวก็เป็นวิธีการคัดเลือกคุณลักษณะที่มีความแม่นยำในการจำแนกประเภทสำหรับปัญหาที่เฉพาะเจาะจง (Janecek, 2009)

สำหรับข้อมูลที่มีจำนวนมิติมาก วิธีฟิลเตอร์มักเป็นวิธีที่ถูกเลือกใช้ในการคัดเลือกคุณลักษณะเนื่องจากประมวลผลได้เร็วกว่าทั้งสองวิธีโดยเฉพาะวิธีฟิลเตอร์แบบคุณลักษณะเดี่ยว แต่วิธีดังกล่าวไม่ได้พิจารณาความสัมพันธ์ระหว่างคุณลักษณะด้วยกันจึงอาจรวมคุณลักษณะที่มีความซ้ำซ้อน หรืออาจละเลยคุณลักษณะที่ไม่มีความสามารถในการจำแนกประเภทถ้าพิจารณาคัดเลือกทีละตัว แต่จะมีความสามารถถ้าอยู่ร่วมกับคุณลักษณะอื่น วิธีฟิลเตอร์แบบหลายคุณลักษณะถึงแม้ว่าจะใช้เวลาในการประมวลผล มากกว่าวิธีฟิลเตอร์แบบคุณลักษณะเดี่ยว แต่ก็ได้นำความสัมพันธ์ระหว่างคุณลักษณะเข้าร่วมในการพิจารณาการคัดเลือกคุณลักษณะ

เมื่อพิจารณาจากความเร็วของวิธีฟิลเตอร์ และความแม่นยำในการจำแนกประเภทของวิธีแรปเปอร์จึงได้มีการนำวิธีทั้งสองมาผสมกันเป็นวิธีใหม่ (Hybrid method) โดยนำข้อดีของแต่ละวิธีมา ซึ่งเป็นแนวโน้มของการคัดเลือกคุณลักษณะในอนาคต

อย่างไรก็ดีไม่มีวิธีการคัดเลือกคุณลักษณะใดที่ดีที่สุดสำหรับทุกสถานการณ์ (Dash & Liu 1997; Zheng & Zhang, 2008; Janecek, 2009) ขึ้นอยู่กับปัจจัยหลายอย่างได้แก่ ลักษณะของข้อมูล (เช่น มีการแจกแจงเป็นแบบเชิงเส้นหรือไม่เชิงเส้น มีตัวแปรปรวนหรือไม่ คุณลักษณะเป็นค่าต่อเนื่องหรือไม่ต่อเนื่อง คุณลักษณะมีความสัมพันธ์กันหรือไม่ เป็นต้น) จำนวนตัวอย่าง และจำนวนคุณลักษณะ ชนิดของขั้นตอนวิธีการเรียนรู้ ลักษณะปัญหา เป็นต้น ดังนั้นจึงควรมีการพัฒนาการคัดเลือกคุณลักษณะให้เหมาะสมกับลักษณะของปัจจัยต่างๆ

เอกสารอ้างอิง

Bramer, M. (2007). *Principles of Data Mining.* Springer.

Dash, M., & Liu, H. (1997). Feature selection for Classification. *Intelligent Data Analysis, 1*, 131-156.

Dash, M., & Liu, H. (2003). Consistency-based search in feature selection. *Artificial Intelligence*, *151*, 155-176.

Forman, G. (2003). An Extensive Empirical Study of Feature Selection Metrics for Text Classification. *Journal of Machine Learning Research, 3*, 1289-1305.

Hall, M. A. (1999). *Correlation-based Feature Subset Selection for Machine Learning.* PhD thesis, Department of Computer Science, University of Waikato, Waikato, N.Z.

Janecek, A. (2009). *Efficient Feature Reduction and Classification Methods: Applications in Drug Discovery and Email Categorization*. PhD dissertation, University of Vienna, Austria.

John, G. H., Kohavi, R., & Pfleger, P. (1994). Irrelevant features and the subset selection problem. In *Machine Learning: Proceedings of the Eleventh International Conference.* Morgan Kaufmann.

Kira, K., & Rendell, L. A. (1992). The feature selection problem: Traditional methods and a new algorithm. In *Proceedings of Ninth National Conference on Artificial Intelligence,* 129–134.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*, 273-324.

Kononenko, I. (1994). Estimating attributes: analysis and extensions of Relief. In: *L.De Raedt and F. Bergadano (eds.): Machine Learning: ECML-94*, 171–182.

Krzanowski, W. J., & Hand, D. J. (2009). A simple method for screening variables before clustering microarray data. *Computational Statistics and Data Analysis, 53*, 2747-2753.

Langley, P., & Sage, S. (1994). Induction of selective Bayesian classifiers. I*n Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence,* 399-406, Seattle, W.A: Morgan Kaufmann.

Liu, H., & Motoda, H. (1998). *Feature Selection for Knowledge Discovery and Data Mining.* MA: Kluwer Academic Publishers Norwell.

Novaković, J., Strbac, P., & Bulatović , D. (2011). Toward optimal feature selection using ranking methods and classification algorithms. *Yugoslav Journal of Operations Research, 21(1),* 119-135.

Pazzani, M. (1995). Searching for dependencies in Bayesian classifiers. In *Proceedings of the Fifth International Workshop on AI and Statistics.*

Vafaie, H., & De Jong, K. (1995). Genetic algorithms as a tool for restructuring feature space representations. In *Proceedings of the International Conference on Tools with A.I.,* IEEE Computer Society Press.

Zheng, H., & Zhang, Y. (2008). Feature selection for high-dimensional data in astronomy. *Advances* in *Space Research, 41,* 1960-1964.