การศึกษาเทคนิคจำแนกประเภทแบบหลายมุมมอง

เปมิกา บุญเสริมส่ง* และศิริสรรพ เหล่าหะเกียรติ

บทคัดย่อ

การแยกแยะแบบหลายมุมมองใช้กับข้อมูลที่มีหลายมุมมองเช่นข้อมูลรูปภาพที่มีด้านหน้า และด้านข้าง โดยเราจะทำการจำแนกประเภทโดยรวมข้อมูลจากมุมมองต่าง ๆ เข้าด้วยกันทำให้ได้ประสิทธิภาพดีขึ้นเมื่อเปรียบเทียบกับการใช้ข้อมูลแบบมุมมองเดียวดังนั้นจึงมีนักวิจัยขยายแนวคิดการแยกแยะแบบหลายมุมมองมาใช้กับข้อมูลแบบมุมมองเดียว โดยแบ่งฟีเจอร์ออกเป็นหลายกลุ่มซึ่งแต่ละกลุ่มจะเป็นตัวแทนของหนึ่งมุมมอง อย่างไรก็ตามยังไม่มีใครเสนอวิธีการแยกแยะข้อมูลออกเป็นหลายมุมมองอย่างเป็นระบบ ในการศึกษานี้เรานำเสนอวิธีการที่เป็นระบบในการสร้างข้อมูลแบบหลายมุมมองขึ้นจากข้อมูลแบบมุมมองเดียว โดยใช้เทคนิคการเปลี่ยนแกนของข้อมูลได้แก่ PCA และ LDA การจำแนกแยกแยะแบบหลายมุมมองที่นำเสนอนี้ใช้เทคนิคการโหวตเสียงส่วนใหญ่ โดยการรวมผลการจำแนกแยกแยะจากตัวจำแนกหลาย ๆตัวที่ถูกเทรนด้วยข้อมูลจากมุมมองที่แตกต่างกัน ผลการทดลองแสดงให้เห็นว่าวิธีการนำเสนอนี้มีประสิทธิภาพมากกว่าวิธีจำแนกแยกแยะแบบมุมมองเดียว

คำสำคัญ: เทคนิคการเรียนรู้ของเครื่อง, การจำแนกแบบหลายมุมมอง, ต้นไม้ตัดสินใจ, พีซีเอ, แอลดีเอ

เปมิกา บุญเสริมส่ง* และศิริสรรพ เหล่าหะเกียรติ

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

*ผู้นิพนธ์ประสานงาน, e-mail: pamika.pay@g.swu.ac.th

A Study of Classification Technique Using Multi-view

Paymika Boonsermsong* and Sirisup Laohakiat

ABSTRACT

Multiview classification is a classification technique used when we have data sets with different views, for example image data sets taken from front and side views. The classification is performed by using data from different views collaboratively, leading to better performance compared with using data from a single view. Recently, researchers have extended the idea to address ordinary data sets with single view by dividing their features to several sets of features, each of which is considered as one view. However, there is no systematic way in dividing the whole features into different views. In this study, we propose a systematic approach in generating Multiview data sets from single view data set using coordinate transformation algorithms including PCA and LDA. The proposed Multiview classification uses majority vote to combine classification results from classifiers trained using data from different views. The experimental results show that the proposed method outperforms single-view classification method.

Keywords: Machine learning, Multiview Classification, Decision Tree, PCA, LDA

Paymika Boonsermsong* and Sirisup Laohakiat

Department of Computer Science, Faculty of Science, Srinakharinwirot University

*Corresponding author, email: pamika.pay@g.swu.ac.th

บทนำ

ในปัจจุบันการนำเทคนิคการเรียนรู้ของเครื่องมาใช้ในการวิเคราะห์ข้อมูลนั้นมีกันอย่างแพร่หลาย แต่นักวิจัยส่วนมากจะใช้วิธีการวิเคราะห์ข้อมูลแบบมุมมองเดียว ซึ่งมีหลายวิธีด้วยกัน เช่น เทคนิคต้นไม้ตัดสินใจ (Decision Tree)[1] เทคนิคการแบ่งกลุ่มข้อมูลแบบเคมีน (K-Means)[2] เทคนิคการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA)[3,4,5,6] เทคนิคการวิเคราะห์จำแนกประเภทเชิงเส้น (Linear Discriminant Analysis: LDA)[6,12] เทคนิคการหาเพื่อนบ้านใกล้เคียงกันมากที่สุด (K-nearest neighbor: K-NN)[7,8] วิธีการถดถอยลอจิสติก (Logistic Regression)[7] เป็นต้น ถึงแม้เทคนิคข้างต้นที่ได้กล่าวมานั้นมีความแม่นยำในระดับหนึ่ง แต่ทว่า ในช่วงหลังมานี้ ได้มีการนำเทคนิคการวิเคราะห์แบบหลายมุมมอง[9] มาช่วยในการแยกแยะข้อมูลเพื่อเพิ่มประสิทธิภาพของการทำงาน

ในบทความนี้ เรานำเสนอแนวความคิดการสร้างข้อมูลแบบหลายมุมมองขึ้นจากข้อมูลแบบมุมมองเดียว ซึ่งเป็นข้อมูลที่ใช้กันในปกติ โดยใช้เทคนิคของการเรียนรู้แบบกลุ่ม[10] มาช่วยเพิ่มประสิทธิภาพ โดยขั้นแรก เราจะสร้างข้อมูลในมุมมองที่แตกต่างกัน จากข้อมูลมุมมองเดียว โดยใช้เทคนิคการบิดแกนข้อมูลชนิดต่างๆ ได้แก่ เทคนิคการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA)[3,4,5,6] และ เทคนิคการวิเคราะห์จำแนกประเภทเชิงเส้น (Linear Discriminant Analysis: LDA)[6,12] จากนั้น เราจะนำข้อมูลแต่ละมุมมองไปสร้างตัวแยกแยะสำหรับแต่ละมุมมองขึ้นมา โดยใช้เทคนิคพื้นฐานที่ใช้กันอยู่ทั่วไป ได้แก่ เทคนิคต้นไม้ตัดสินใจ หรือ เทคนิคการหาเพื่อนบ้านใกล้เคียงกันมากที่สุด เมื่อได้ตัวแยกแยะสำหรับแต่ละมุมมองที่แตกต่างกันแล้ว เราจะรวมคำตอบของตัวแยกแยะจากทุกมุมมองด้วยการใช้เทคนิคการเรียนรู้แบบกลุ่ม

โดยเทคนิคการเรียนรูปแบบกลุ่มนี้ ได้มีงานที่ศึกษามาก่อนหน้านี้ เช่น ปรเมษฐ์ ธันวานนท์และคณะ[10] ได้ใช้เทคนิคการเรียนรู้แบบรวมกลุ่มเพื่อเพิ่มประสิทธิภาพในการพยากรณ์ โดยใช้เทคนิคการจำแนกประเภทแบบการสุ่มป่าไม้ (Random Forest)[9] เทคนิคการจำแนกประเภทแบบซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines: SVM)[8] เทคนิคการหาเพื่อนบ้านใกล้เคียงกันมากที่สุด (K-nearest neighbor: K-NN)[7,8] และเทคนิค Naïve Bayes[13] ในการทำนายราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย โดยเมื่อใช้เทคนิคการเรียนรู้แบบรวมกลุ่ม (Ensemble Model)[10] สามารถเพิ่มประสิทธิภาพในการพยากรณ์ได้ 5-14% และเพิ่มผลตอบแทนในการลงทุน 1-3% และปัจจัยหลักที่สำคัญที่ทำให้ประสิทธิภาพเพิ่มขึ้นคือ ข้อมูลดัชนีชี้วัดทางเทคนิค และจำนวนวันที่ถือหุ้น เดช ธรรมศิริ และพยุงมีสัจ[11] ได้นำเสนอการจำแนกข้อมูลด้วยวิธีแบบร่วมกันตัดสินใจจากพื้นฐานของเทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนร่วมกับการเลือกตัวแทนที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรม โดยใช้ข้อมูลAustralian Credit ผลที่ได้ใช้เทคนิคต้นไม้ตัดสินใจ 3 โมเดล และเทคนิคโครงข่ายประสาทเทียม 2 โมเดลผลลัพธ์ความถูกต้องเท่ากับ 89.01% ส่วนข้อมูล German Credit ได้ใช้เทคนิคต้นไม้ตัดสินใจ 6 โมเดล และเทคนิคโครงข่ายประสาทเทียม 5 โมเดล ได้ลัพธ์ความถูกต้องเท่ากับ 71.50% และสุดท้ายข้อมูล Bankruptcy Data ใช้การจำแนกข้อมูลทั้งหมด 10 โมเดลร่วมกัน ซึ่งประกอบด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน 4 โมเดล เทคนิคต้นไม้ตัดสินใจ 4 โมเดล และเทคนิคโครงข่ายประสาทเทียม 2 โมเดล ผลลัพธ์ความถูกต้องเท่ากับ 71.50% จากผลที่ได้นั้นทำให้เห็นว่าการใช้เทคนิคร่วมกันตัดสินใจสามารถช่วยเพิ่มประสิทธิภาพความแม่นยำในการจำแนกข้อมูลได้มากกว่าการใช้มุมมองเดียว

อุปกรณ์และวิธีทดลอง

Dell Inspiron 14 7460 Intel Core i7-7500 CPU 2.70GHz Ram 8 GB 64bit OS Windows 10 และใช้ Python 3.0 บน Jupyter Notebook version 5.7.4

ข้อมูลที่นำมาทดลอง ใช้ข้อมูลจาก UCI โดยใช้ทั้งหมด 4 ข้อมูลดังนี้

1. Image Segmentation เป็นข้อมูลการแบ่งกลุ่มภาพที่อยู่กลางแจ้ง ประกอบด้วยข้อมูลทั้งสิ้น 2310 ข้อมูล โดยแต่ละข้อมูลได้คุณสมบัติของรูปภาพขนาดกริด 3x3 โดยข้อมูลแต่ละระเบียน จะเป็นข้อมูลภาพ จะอยู่ในกลุ่มใดกลุมหนึ่งในเจ็ดกลุ่มดังนี้ ภาพพื้นทางเท้า, ภาพท้องฟ้า, ใบไม้, พื้นถนน, หน้าต่าง, ทางเดิน, และต้นหญ้า

2. Spambase เป็นข้อมูลจดหมายขยะโดยจำแนกว่าจดหมายที่ได้รับเป็นจดหมายขยะหรือไม่

3. Ionosphese เป็นข้อมูลชุดเสาอากาศความถี่สูงจำนวน 16 ส่วน เพื่อเปรียบเทียบในชั้นบรรยากาศ ว่าสัญญาณผ่านชั้นบรรยากาศจากเรดาร์ "ดี" หรือ "ไม่ดี"

4. Breast Cancer Wisconsin เป็นชุดข้อมูลเพื่อนำมาวิเคราะห์ว่าเป็นมะเร็งเต้านมหรือไม่

วิธีทดลอง

จากข้อมูลทั้ง 4 ข้อมูลที่นำมาจาก UCI ผู้วิจัยใช้เทคนิคการจำแนกประเภทของข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree)[1] และ เทคนิคการหาเพื่อนบ้านใกล้เคียงกันมากที่สุด (K-nearest neighbor: K-NN)[7,8] นำมาสร้างมุมมองด้วยวิธีการที่แตกต่างกัน สามมุมมอง โดยในมุมมองแรก จะใช้ข้อมูลตั้งต้นเดิม โดยใช้ฟีเจอร์ทั้งหมด ข้อมูลในมุมมองที่สอง จะสร้างโดยใช้เทคนิค PCA บิดแกนไปในทิศทางที่ข้อมูลอยู่มากที่สุดโดยแกนแรกชี้ไปในทิศที่ข้อมูลอยู่มากที่สุด แกนที่สองจะต้องตั้งฉากกับแกนแรกเสมอและจะมีข้อมูลกระจายตัวลองลงมาจากแกนแรก เป็นเช่นนี้ไปตามลำดับ ในงานวิจัยนี้ ผู้วิจัยได้ทำ PCA ก่อน จากนั้นจะเลือกแกนหลักเพียง 7 แกนหลัก (n_components=7) แล้วจึงไปทำสร้างตัวแยกแยะแบบ Decision Tree ส่วนในมุมมองที่สามจะใช้เทคนิค LDA ในการเปลี่นแปลงแกนข้อมูล ทั้งนี้เนื่องจากในแต่ละชุดข้อมูล จำนวนของมิติที่เหลือในเทคนิค LDA นี้ จะเท่ากับจำนวนของคลาสของข้อมูลลบด้วยหนึ่ง ดังนั้นในมุมมองนี้ จำนวนมิติของข้อมูลจึงถูกกำหนดด้วยจำนวนคลาสของข้อมูลอยู่แล้ว หลังจากได้ตัวแยกแยะสำหรับทั้งสามมุมมองแล้ว เราจะหาคลาสของคำตอบรวมจากทั้งสามตัวแยกแยะโดยใช้ใช้วิธี Majority Vote[10,12] ในการทดลองนี้ ผู้ทดลองได้ทดลองเปลี่ยนเทคนิคการเรียนรูปแบบต่างๆ ได้แก่ เทคนิคต้นไม้ตัดสินใจ หรือ เทคนิคการหาเพื่อนบ้านใกล้เคียงกันมากที่สุด เพื่อค้นหาวิธีที่ทำให้ได้รับประสิทธิภาพในการทำงานสูงสุด

ผลการทดลอง

จากการทดลองแรก เราสร้างตัวแยกแยะสำหรับข้อมูลแต่ละมุมมองด้วยด้วยเทคนิคต้นไม้ตัดสินใจ ผลลัพธ์ที่ได้แสดงในตารางที่ 1 โดย Full Feature หมายถึง การนำข้อมูลเริ่มต้นมาใช้สร้างตัวแยกแยะ PCA หมายถึง การนำข้อมูลที่ได้จากการใช้เทคนิคการลดมิติด้วย PCA มาสร้างตัวแยกแยะ และ LDA หมายถึง การนำข้อมูลที่ได้จากการใช้เทคนิคการลดมิติด้วย LDA มาใช้สร้างตัวแยกแยะ

ตารางที่ 1 ตารางการทดสอบแบบ Decision Tree

DATA	Full Feature	PCA	LDA	Multi View
Image Segmentation	0.94	0.91	0.92	0.95
Spambase	0.90	0.85	0.86	0.92
Ionosphese	0.86	0.86	0.86	0.91
Breast Cancer Wisconsin	0.94	0.90	0.97	0.97

จากตารางที่ 1 จะเห็นได้ว่าค่า ความแม่นยำ ของการใช้ ตัวแยกแยะที่ใช้การรวมค่าจากหลายมุมมอง ได้ผลดีกว่าการแยกแยะโดยใช้ข้อมูลจากแต่ละมุมมอง ได้แก่ การนำฟีเจอร์ทั้งหมดมาใช้ การใช้เทคนิค PCA และ การใช้เทคนิค LDA มาทำการทดสอบ

จากการทดลองที่สอง เราสร้างตัวแยกแยะสำหรับข้อมูลแต่ละมุมมองด้วยด้วยเทคนิคการหาเพื่อนบ้านใกล้เคียงกันมากที่สุด ผลลัพธ์ที่ได้แสดงในตารางที่ 1 โดย Full Feature หมายถึง การนำข้อมูลเริ่มต้นมาใช้สร้างตัวแยกแยะ PCA หมายถึง การนำข้อมูลที่ได้จากการใช้เทคนิคการลดมิติด้วย PCA มาสร้างตัวแยกแยะ และ LDA หมายถึง การนำข้อมูลที่ได้จากการใช้เทคนิคการลดมิติด้วย LDA มาใช้สร้างตัวแยกแยะ

ตารางที่ 2 ตารางการทดสอบแบบ K-NN

DATA	Full Feature	PCA	LDA	Multi View
Image Segmentation	0.91	0.91	0.95	0.91
Spambase	0.78	0.77	0.90	0.80
Ionosphese	0.83	0.87	0.89	0.87
Breast Cancer Wisconsin	0.68	0.68	0.98	0.68

จากตารางที่ 2 จะเห็นได้ว่าค่าของ Accuracy ของการใช้ LDA ได้ผลดีกว่าในกรณีอื่นทั้งหมด รวมทั้งยังดีกว่าในกรณีของการใช้ข้อมูลแบบหลายมุมมองอีกด้วย ทั้งนี้ เนื่องจาก ผลลัพธ์ที่ได้จากการใช้เทคนิค PCA หรือ การใช้ข้อมูลตั้งต้น ต่ำกว่าการใช้เทคนิค LDA ต่ำกว่าการใช้เทคนิค LDA มาก เมื่อนำผลลัพธ์จากเทคนิคทั้งสองมาร่วมทำการโหวตแล้ว ก็ยังไม่อาจทำให้ดีไปกว่า ผลของการใช้เทคนิค LDA แต่เพียงอย่างเดียว แต่ทั้งนี้ มีข้อสังเกตว่า ค่าความแม่นยำที่ได้จากเทคนิคการใช้ข้อมูลแบบหลายมุมมอง ก็ยังมีค่าดีกว่า การใช้เทคนิคมุมมองเดียวในกรณีของ Full Feature หรือ PCA

ในการทดลองสุดท้าย เราเปรียบเทียบความแม่นยำระหว่าง การใช้เทคนิคการแยกแยะแบบหลายมุมมอง กับการใช้เทคนิคการแยกแยะแบบหลายโมเดล โดยโมเดลที่นำมาใช้ทดลองได้แก่ โมเดล logistic regression, random forest และ K-NN แสดงในตารางที่ 3

ตารางที่ 3 ตารางเปรียบเทียบ Multi View กับ Multiple Model

DATA	Multi View Decision Tree	Multi View KNN	Multiple Model
Image Segmentation	0.95	0.91	0.93
Spambase	0.92	0.82	0.91
Ionosphese	0.91	0.87	0.87
Breast Cancer Wisconsin	0.97	0.68	0.66

จากตารางที่ 3 จะเห็นได้ว่าค่าของ Accuracy ของการใช้เทคนิคการแยกแยะแบบหลายมุมมอง โดยใช้ตัวแยกแยะ แบบ Decision Tree ให้ค่าความแม่นยำที่สูงที่สุด

สรุปและวิจารณ์ผลการทดลอง

จากผลการทดลองจึงสรุปได้ว่าการใช้การแยกแยะแบบหลายมุมมอง โดยใช้ตัวแยกแยะ แบบต้นไม้ตัดสินใจจะให้ค่าความแม่นยำ สูงกว่า การใช้เทคนิคแบบอื่น ไม่ว่า จะเป็นการแยกแยะแบบมุมมองเดียว หรือ การแยกแยะโดยใช้หลายแบบจำลอง แต่ทั้งนี้ การใช้เทคนิคการแยกแยะแบบหลายมุมมองนี้ ยังขึ้นกับชนิดของตัวแยกแยะที่นำมาใช้

เป็นอย่างมาก เมื่อเราสร้างตัวแยกแยะการหาเพื่อนบ้านใกล้เคียงกันมากที่สุด จะได้ผลที่ต่ำกว่าการใช้ต้นไม้ตัดสินใจเป็นอย่างมาก ดังนั้น จึงต้องมีการศึกษาเพิ่มเติมถึงสาเหตุที่ทำให้ประสิทธิภาพการทำงานของทั้งสองกรณีมีความแตกต่างกันเช่นนี้ ต่อไป

กิตติกรรมประกาศ

ขอขอบพระคุณภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ สำหรับสถานที่ในการทำวิจัยครั้งนี้

เอกสารอ้างอิง

[1] มานิตา สองสี และคณะ. (2013). เหมืองข้อมูลเพื่อการทำนายสถานภาพของนักศึกษาวิทยาลัยเทคโนโลยีภาคใต้. *งานประชุมวิชาการระดับนานาชาติและระดับชาติวิศวศึกษา ครั้งที่ 11*, 9-11 พฤษภาคม 2556 โรงแรมภูเก็ตเกรซแลนด์ รีสอร์ท แอนด์ สปา. คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์: วิศวศึกษา

[2] สุพจน์ เฮงพระพรหม. (2551). *กลุ่มก้อนตัวจำแนกประเภทกำหนดการพันธุกรรมสำหรับข้อมูลไมโครอาร์เรย์.* กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

[3] วิทวัส วิทยาไกรเลิศ. (2556). *การวิเคราะห์ภาพใบหน้า 3 มิติสำหรับการระบุตัวตนและยืนยัน ตัวบุคคล.* นครราชสีมา: มหาวิทยาลัยเทคโนโลยีสุรนารี.

[4] สุเมธ พิลึก. (2559). การวิเคราะห์องค์ประกอบสมรรถภาพทางวิชาชีพของผู้สำเร็จการศึกษาสาขาวิชาคอมพิวเตอร์ธุรกิจตามความต้องการของตลาดแรงงาน. *สมาคมสถาบันอุดมศึกษาเอกชนแห่งประเทศไทย ในพระราชูปถัมภ์ สมเด็จพระเทพรัตนราชสุดาฯ สยามบรมราชกุมารี*. 22(1), 96-109.

[5] ธีรยุทธ์ เมืองแก้ว. (2554). *เรื่อง การวิเคราะห์องค์ประกอบของแบบประเมินในการประหยัดพลังงานไฟฟ้า.* กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.

[6] ธนพล พุ่มลำเจียก. (2559). *เรื่อง การรู้จำอาจมณ์ใบหน้าจากวีดีโอ โดยใช้ตัวกรองกาบอร์ วิธีการวิเคราะห์องค์ประกอบหลักและการวิเคราะห์จำแนกประเภทเชิงเส้น.* กรุงเทพฯ: สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง

[7] สุรวัชร ศรีเปารยะ และสายชล สินสมบูรณ์ทอง. (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง : กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดียว. วารสารวิทยาศาสตร์และเทคโนโลยี. 25(5), 839-853.

[8] ภูริพัทร์ ทองคำ. (2559). *เรื่อง อัลกอริทึมแบบรวมสำหรับการเลือกคุณสมบัติของข้อมูล.* กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์

[9] Jing Zhao, Xijiong Xie, Xin Xu, Shiliang Sun. (2017). Multi-view learning overview: Recent progress and new challenges. *Journal on Multi-Sensor, Multi-Source Information Fusion*, 38(2017), 43-54.

[10] ปรเมษฐ์ ธันวานนท์, ชัยกร ยิ่งเสร, และ วรพล พงษ์เพ็ชร. (2560). การประยุกต์ใช้โมเดลการเรียนรู้แบบรวมกลุ่มเพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย. *วารสารวิทยาการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีมหานคร*. 7(1), 12-21.

[11] เดช ธรรมศิริ และ พยุง มีสัจ. (2553). การจำแนกข้อมูลด้วยวิธีแบบร่วมกันตัดสินใจจากพื้นฐานของเทคนิคต้นไม้ตัดสินใจเทคนิคโครงข่ายประสาทเทียม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนร่วมกับการเลือกตัวแทนที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรม. วารสารวิชาการพระจอมเกล้าพระนครเหนือ. 21(2), 293-303.

[12] Kumar R, Goyal M.K, Ahmed P, Kumar A. (2012). Unconstrained Handwritten Numeral Recognition Using Majority Voting Classifier. *IEEE International Conference on Parallel, Distributed and Grid Computing*, 978 (2), 284-289.

[13] นพมาศ ปักเข็ม, ชนิดา จันมณีย, และ ศิวกร อุยสุย. (2560). การจำแนกประเภทภูมิปัญญาท้องถิ่นของไทยแบบอัตโนมัติโดยวิธีการทางเหมืองข้อมูล. *วารสารมหาวิทยาลัยทักษิณ*, 20(3), 300-307.