

การจำแนกชนิดของพืชด้วยวิธีเพื่อนบ้านใกล้สุดร่วมกับการเลือกตัวแทนที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรมโดยใช้คุณลักษณะรูปร่างและพื้นผิวของใบพืช*

Classification of plants with the technique of k-Nearest Neighbor (k-NN) and the selection of appropriate agents with Genetic Algorithm (GA) using the features of shapes and surfaces of leaves

ธวัช รวมทรัพย์(Tavat Ruamsub)**

สุรศักดิ์ มั่งสิงห์(Surasak Mungsing)**

บทคัดย่อ

เทคนิคในการจำแนกชนิดของใบพืชมีหลายเทคนิค เช่น C4.5, Naive Bayes, k-Nearest Neighbor (k-NN) , Genetic Algorithm (GA) และ Support Vector Machine (SVM) เป็นต้น ซึ่งแต่ละเทคนิคก็มีข้อดีและข้อเสียที่แตกต่างกันไปขึ้นอยู่กับปัญหาและวัตถุประสงค์การนำไปใช้ในการหาผลลัพธ์ ด้วยคุณลักษณะของใบพืชซึ่งประกอบไปด้วย รูปใบ เส้นใบ ปลายใบ โคนใบ ขอบใบ และขนาดของใบ ผู้วิจัยพบว่า สามารถนำคุณลักษณะต่างๆของใบพืช มาใช้ในการจำแนกชนิดของพืชได้ งานวิจัยครั้งนี้เป็นการนำเสนอวิธีการจำแนกชนิดของพืช โดยใช้วิธี เลือกตัวแทนที่เหมาะสม ด้วยขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm) และใช้เทคนิคเพื่อนบ้านใกล้สุด (k-nearest Neighbor) ในการจำแนกชนิดของพืช สำหรับกลุ่มตัวอย่างที่ใช้ในการทดลองประกอบด้วย ใบพืช 30 สายพันธุ์ จำนวน 340 ใบ จากผลการทดลองโดยใช้เทคนิคเพื่อนบ้านใกล้สุดเพียงอย่างเดียวให้ผลลัพธ์ความถูกต้อง (accuracy) 79.12% ค่าความครบถ้วน (recall) 79.50% และค่าความแม่นยำ (precision) 73.17% และเมื่อใช้การเลือกตัวแทนที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรมร่วมกับเทคนิคเพื่อนบ้านใกล้สุดให้ผลลัพธ์ความถูกต้องในการจำแนกชนิดของพืช (accuracy) 86.27% ค่าความครบถ้วน (recall) 83.06% และค่าความแม่นยำ (precision) 84.06%

คำสำคัญ: การจำแนกประเภท ขั้นตอนวิธีเชิงพันธุกรรม ขั้นตอนวิธีเพื่อนบ้านใกล้สุด

* บทความวิจัยนี้ผู้วิจัยมีวัตถุประสงค์เพื่อใช้ขั้นตอนวิธีเชิงพันธุกรรมในการเลือกคุณลักษณะที่เหมาะสมของใบพืชและใช้ขั้นตอนวิธีเพื่อนบ้านใกล้สุดในการจำแนกชนิดพืชเพื่อให้การจำแนกชนิดพืชด้วยขั้นตอนวิธีเพื่อนบ้านใกล้สุดมีประสิทธิภาพเพิ่มขึ้นในการจำแนกชนิดพืช

** หลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม 61 ถนนพหลโยธิน แขวงเสนานิคม เขตจตุจักร กรุงเทพมหานคร 10900 E-mail: tavat69@gmail.com

Doctor of Philosophy in Information Technology. Faculty of Information Technology, Sripatum University.
61 Paholyothin Rd., Senanikom Subdistrict, Jatujak District, Bangkok 10900 E-mail: tavat69@gmail.com

** ผู้ช่วยศาสตราจารย์ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม

Assistant Professor, Faculty of Information Technology, Sripatum University.

Abstract

There are many techniques in the classification of plant leaves such as C4.5, Naive Bayes, k-Nearest Neighbor (k-NN), Genetic Algorithm (GA), Support Vector Machine (SVM), etc. Each technique has its own different advantages and disadvantages depending on problems and objectives of implementation to obtain the results. For the features of plant leaves, which consist of leaf shapes, leaf veins, leaf apexes, leaf bases, leaf margins, and leaf sizes, the researcher found that the various features of plant leaves could be used in the classification of plants. This research presented the method of the classification of plants by using the selection of appropriate agents with Genetic Algorithm (GA) and with the technique of k-Nearest Neighbor (k-NN). The samples used in the experiment consisted of 30 species of 340 plant leaves. From the result of the experiment only with the technique of k-Nearest Neighbor (k-NN), the accuracy in the classification of plants was by 79.12%, the recall was by 79.50% and the precision was by 73.17% and with the selection of appropriate agents with Genetic Algorithm (GA) as well as the technique of k-Nearest Neighbor (k-NN), the accuracy in the classification of plants was by 86.27%, the recall was by 83.06% and the precision was by 84.06%.

Keywords: Classification, Genetic Algorithm, k-Nearest Neighbor (k-NN)

บทนำ

สายพันธุ์ของพืชที่มีการสำรวจในโลกนี้ที่ได้รับการตั้งชื่อมีประมาณ 250,000 – 270,000 สายพันธุ์ (Guo et al, 2004) ในการที่จะระบุชนิดของพันธุ์พืชให้ถูกต้องจากการที่ได้พบเห็นในที่ต่างๆจึงเป็นเรื่องที่ยากมากเนื่องจากสายพันธุ์พืชมีมากมาย ปัจจุบันอุปกรณ์เทคโนโลยีและการสื่อสารมีความทันสมัยมาก การนำเทคโนโลยีเข้ามาใช้ช่วยเหลือในการจำแนกพันธุ์พืช จึงช่วยย่นระยะเวลา และมีความแม่นยำในการระบุชนิดพันธุ์พืชได้

ขั้นตอนวิธีเพื่อนบ้านใกล้สุด (k-Nearest Neighbors) (Daniel T. Larose, 2005; Xindong Wu et al, 2008) เป็นวิธีการในการจัดแบ่ง คลาส โดยจะตัดสินใจว่าคลาสใดที่จะแทนเงื่อนไข หรือกรณีใหม่ ๆ ได้บ้าง เป็นวิธีการหนึ่งสำหรับ แก้ปัญหาประมาณค่าฟังก์ชันนอนพาราเมตริก สำหรับ การจำแนกกลุ่มของข้อมูลที่ไม่เป็นรูปร่างที่ดี หรือ ข้อมูลที่กระจัดกระจาย โดยทำการตรวจสอบจำนวน บางจำนวนของกรณีหรือเงื่อนไขที่เหมือนกัน หรือ ใกล้เคียงกันมากที่สุด เท่ากับจำนวน k ที่ต้องการ โดยการหาระยะทางที่ใกล้สุด

ขั้นตอนวิธีทางพันธุกรรม (Genetic Algorithm : GA) ใช้หลักการมาจากทฤษฎีการวิวัฒนาการของ Charles Darwin ว่าด้วยสิ่งมีชีวิตที่แข็งแรงที่สุดจะมีโอกาสสืบทอดสายพันธุ์ต่อไป โดยทาง (John H. Holland, 1975) ได้นำมาเสนอหลักการ และจากนั้น (David E. Goldberg, 1989) ได้นำ GA มาปรับปรุงและประยุกต์ใช้กับงานทางวิศวกรรมเป็นผลให้ GA เป็นที่รู้จัก และนำมาประยุกต์ใช้ในงานกันอย่างแพร่หลาย ปัจจุบัน GA ได้มี

การปรับปรุงเพื่อเพิ่มประสิทธิภาพการค้นหาให้มีความสามารถ และประสิทธิภาพสูงขึ้น (ดร.อาณัติ ลีมีคเดช, 2546)

ขั้นตอนการทำงานทั่วไปของเจเนติก อัลกอริทึม สามารถเขียนเป็นฟังก์ชันการทำงานได้ดังนี้ (Dr. R.K Bhattacharjya, 2012)

```
functionGeneticAlgo()  
{  
    Initialize population;  
    fitness function;  
    While(fitness value != termination criteria)  
    {  
        SelectionCrossover;  
        Mutation;Calculate  
        fitnessfunction;  
    }  
}
```

งานวิจัยนี้ได้ใช้คุณลักษณะของใบพืชซึ่งประกอบไปด้วย คุณลักษณะรูปทรง และคุณลักษณะพื้นผิว มาใช้เป็นองค์ประกอบในการจำแนกชนิดของพืช โดยการใช้ขั้นตอนวิธีเชิงพันธุกรรมสำหรับเลือกตัวแทนที่เหมาะสมเพื่อใช้ในการจำแนกชนิดของพืช และใช้ขั้นตอนวิธีเพื่อนบ้านใกล้สุด (k-Nearest Neighbor) สำหรับการจำแนกชนิดของพืช

งานวิจัยนี้มีวัตถุประสงค์ 2 ประการ คือ

1. เพื่อนำกระบวนการขั้นตอนวิธีเชิงพันธุกรรมแต่ละขั้นตอนเลือกตัวแทนที่เหมาะสมสำหรับนำไปใช้ในการจำแนกชนิดของพืช
2. เพื่อใช้ขั้นตอนวิธีเพื่อนบ้านใกล้สุด สำหรับการจำแนกชนิดพืชโดยใช้คุณลักษณะรูปทรงและพื้นผิวของใบพืช

งานวิจัยที่เกี่ยวข้อง

งานวิจัยครั้งนี้ผู้วิจัยได้ทำการค้นคว้าข้อมูลจากเอกสาร งานวิจัย และผลงานวิชาการต่างๆ ที่เกี่ยวข้องในการจำแนกใบพืช มีดังนี้

งานวิจัยที่เกี่ยวข้องทางด้านเทคนิคการจำแนก (Prof Meeta K et al.) ได้สำรวจเทคนิคสำหรับการจำแนกใบพืชซึ่งมีหลายเทคนิค เช่น k-Nearest Neighbor Classification (k-NN) , Probabilistic Neural Network(PNN), Genetic Algorithm(GA), Support Vector Machine (SVM), และ Principal Component Analysis (PCA) ซึ่งแต่ละเทคนิคก็มีข้อดีและข้อเสียที่แตกต่างกันไปขึ้นอยู่กับปัญหาและวิธีการหาผลลัพธ์

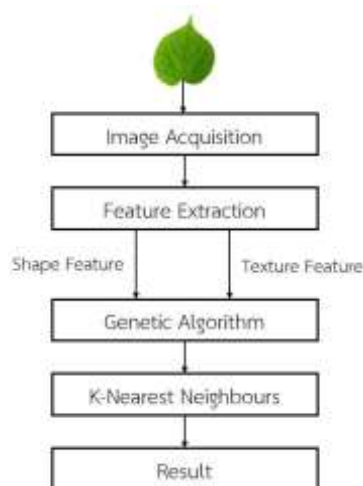
งานวิจัยที่เกี่ยวข้องทางด้านการจำแนกชนิดพืชทางด้านเทคนิคต่างๆมีหลายเทคนิคเช่น โครงข่ายประสาทเทียม(Artificial Neural Network) (Hong and Chi, 2003;2006) ได้ประยุกต์ใช้ Artificial neural network (ANN) โดยการใช้การสกัดลายใบพืชสำหรับใช้ในการรู้จำ (Jiazhi and He, 2008) ได้ประยุกต์ใช้ Artificial neural network (ANN) ใช้ในการรู้จำ โดยการใช้การสกัดใบพืชออกเป็น ส่วน (Stephen et al, 2007) ได้ประยุกต์ใช้ Probabilistic neural network (PNN) ใช้ในการรู้จำ โดยสกัดคุณลักษณะของใบพืชออกเป็น 12 คุณลักษณะ (Huang and Peng, 2008) ได้ประยุกต์ใช้ Probabilistic neural network (PNN) ใช้ในการรู้จำ โดยใช้คุณสมบัติรูปร่างและพื้นผิวของใบพืช (Yun et al, 2005) ได้ประยุกต์ใช้ Cellular neural network (CNN) โดยการใช้การสกัดลายใบพืช (Panagiotis et al, 2005) ได้ประยุกต์ใช้ feed-forward neural network โดยสกัดคุณลักษณะของใบพืช (Xiao et al, 2005) ได้ประยุกต์ใช้ k-Nearest Neighbor (kNN) ร่วมกับ Probabilistic neural network (PNN) ใช้ในการรู้จำใบพืช

งานวิจัยที่เกี่ยวข้องทางด้าน GA (Padmavathi, 2015) ได้ประยุกต์ใช้ GA ตรวจสอบใบพืชที่เป็นโรคหรือติดเชื้อ โดยการหาค่าบริเวณใบพืชที่เป็นโรคและบริเวณที่ไม่เป็นโรค (Tsai et al, 2013; Yi et al, 2012; Kabir et al, 2011) ได้ประยุกต์ใช้ GA เพื่อเลือกคุณลักษณะที่เหมาะสมของข้อมูล (Feature Selection) และยังมีผู้วิจัยหลายท่านดังเช่น (Ilhan and Tezel, 2013; Huang et al, 2011; Chou et al, 2010) ได้นำ GA มาประยุกต์ใช้สำหรับการหาค่าพารามิเตอร์ ที่เหมาะสมสำหรับตัวจำแนกข้อมูล (Pei-Yong et al, 2009) ได้ใช้ K-NN และ Bayes เป็นตัวจำแนกข้อมูลพร้อมด้วยการใช้วิธีการเรียนรู้แบบรวมกลุ่ม และใช้วิธีเชิงพันธุกรรม ช่วยเลือกคุณลักษณะที่เหมาะสม

วิธีการวิจัย

งานวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการวิจัยประกอบด้วย 5 ขั้นตอนดังนี้

1. การเก็บและเตรียมภาพใบพืช
2. การแยกคุณลักษณะ
3. วิธีเชิงพันธุกรรม
4. การจำแนกภาพใบพืชด้วยขั้นตอนวิธีเพื่อนบ้านใกล้เคียง
5. การวัดประสิทธิภาพ ดังภาพที่ 1



ภาพที่ 1 ขั้นตอนวิธีดำเนินการวิจัย

ขั้นตอนที่ 1. การเก็บและเตรียมภาพใบพืช (Image Acquisition)

สำหรับข้อมูลที่นำมาใช้ในการทดลองการจำแนกใบพืช ซึ่งภาพใบพืชแต่ละใบถูกถ่ายด้วยกล้องดิจิทัลโดยกำหนดค่าความละเอียดที่ 720 x 920 pixel และจัดเก็บให้อยู่ในรูปแบบนามสกุล JPEG ภาพใบพืชที่ใช้เป็นกลุ่มตัวอย่างมี 30 สายพันธุ์ รวมทั้งหมด 340 ใบ ผู้วิจัยได้ใช้ข้อมูลใบพืช “Leaf” dataset สำหรับใช้เป็นข้อมูลในการทดลอง (Pedro F.B.Silva et al, 2014)



ภาพที่ 2 สายพันธุ์พืช 30 สายพันธุ์

ขั้นตอนที่ 2. การแยกคุณลักษณะ (Feature Extraction)

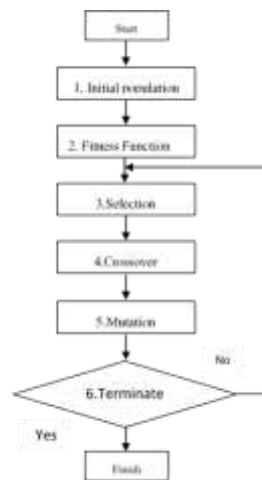
เนื่องจากลักษณะรูปร่าง และองค์ประกอบต่างๆของใบพืชมีคุณลักษณะที่แตกต่างกัน และคล้ายคลึงกันจึงต้องแยกคุณลักษณะต่างๆของภาพใบพืชออกเป็น 2 กลุ่ม กลุ่มที่1 ใช้คุณลักษณะรูปร่างของใบพืช จำนวน 8 คุณลักษณะ กลุ่มที่2 ใช้คุณลักษณะของพื้นผิวของใบพืช จำนวน 6 คุณลักษณะ รวมทั้งสิ้น 14 คุณลักษณะ ดังภาพที่ 3

คุณลักษณะ	คุณลักษณะรูปร่าง	คุณลักษณะ	คุณลักษณะพื้นผิว
1	Eccentricity	1	Average Intensity
2	Aspect Ratio	2	Average Contrast
3	Elongation	3	Smoothness
4	Solidity	4	Third moment
5	Stochastic Convexity	5	Uniformity
6	Isoperimetric Factor	6	Entropy
7	Maximal Indentation Depth		
8	Lobedness		

ภาพที่ 3 คุณลักษณะรูปร่างและคุณลักษณะพื้นผิว

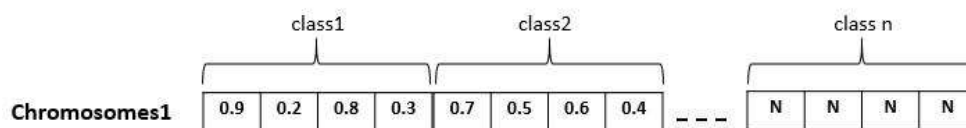
ขั้นตอนที่ 3. ขั้นตอนวิธีเชิงพันธุกรรม

ขั้นตอนวิธีเชิงพันธุกรรม ซึ่งมีขั้นตอนดังนี้ (Z.Michalewicz, 1972)



ภาพที่ 4 ขั้นตอนการทำงานของขั้นตอนวิธีเชิงพันธุกรรม

3.1 การสร้างประชากรเริ่มต้น (initial population) การสร้างประชากรเริ่มต้นโดยใช้การสุ่มค่าให้ยีน ระบบจะแปลง (encode) ผลลัพธ์ที่เป็นไปได้ของปัญหาคณิตศาสตร์ให้อยู่ในรูปโครโมโซม ตัวแปรที่ต้องการค้นหา หมายถึง ยีน (gene) หนึ่งยีนเมื่อนำมาประกอบกันจึงได้โครโมโซมหรือรหัสพันธุกรรม (genotype)



ภาพที่ 5 ตัวอย่างการกำหนด initial population

เมื่อกำหนดวิธีการแปลงแล้ว ระบบจะสุ่มกลุ่มโครโมโซมนี้ขึ้นมาเป็นประชากร จำนวนโครโมโซมในประชากรแต่ละรุ่นขึ้นกับค่าพารามิเตอร์ที่กำหนด ความยาวของโครโมโซมขึ้นอยู่กับจำนวนตัวแปรที่ต้องการหาค่าในระบบ และมีผลต่อการกำหนดพารามิเตอร์จำนวนประชากร เพื่อหาคำตอบที่ดีที่สุด

3.2 การประเมินค่าความเหมาะสม (Fitness Function) เป็นฟังก์ชันที่มีความสำคัญเนื่องจาก Fitness Function จะให้ค่า fitness value กับแต่ละโครโมโซม เพื่อใช้ในการตัดสินใจในการเลือก (selection) โครโมโซม parents หรือโครโมโซมที่จะกลายเป็นรุ่นต่อไป โดยถ้าหากโครโมโซมใดมีค่า fitness value ต่ำก็จะถูกทิ้งไปเปิดโอกาสให้โครโมโซมที่มีค่า fitness value สูงได้สืบทอดเป็นรุ่นต่อไป ดังสมการที่ 1

$$UD_{chromosome_i} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

p_i คือ ค่าที่อยู่ในยีนของ Chromosome_i

q_i คือ ค่าที่อยู่ในยีนของ Data train_n

n คือ จำนวน Data train

3.3. การคัดเลือกโครโมโซม (selection) เป็นขั้นตอนในการจับคู่โครโมโซม parents เพื่อเป็นประชากรรุ่นต่อไป โดยใช้วงล้อเสี่ยงทาย (Roulette wheel selection) โครโมโซมที่มีค่าความเหมาะสมที่ดีกว่ามีโอกาสถูกเลือกมากกว่า ดังสมการที่ 2

$$r = \frac{1 - f(x)}{\sum(1 - f(x))} \quad (2)$$

$f(x)$ คือ ข้อมูลจริงที่เกิดขึ้น

r คือ ข้อมูลที่เกิดจากการพยากรณ์

ทำการคำนวณค่าฟังก์ชัน $f(x)$ โดยทำการคำนวณค่าความเหมาะสม

3.4. การครอสโอเวอร์ (crossover) คือการสลับค่าของ โครโมโซมระหว่างประชากรที่ได้คัดเลือกขึ้นอยู่กับค่าความน่าจะเป็นในการแลกเปลี่ยนยีน ความน่าจะเป็นในการสลับสายพันธุ์ (Probabilities of crossover: Pc) วิธี crossover มีหลายวิธีงานวิจัยนี้เลือกใช้วิธี Arithmetic crossover (AC) ดังสมการที่ 3

$$Child_1 = r * parent1 + (1 - r) * parent2 \quad Child_2 = (1 - r) * parent1 + r * parent2 \quad (3)$$

เมื่อ r คือค่าที่สุ่มอยู่ในช่วง $[0,1]$

3.5. การกลายพันธุ์ (mutation) การกลายพันธุ์เป็นกระบวนการสุ่ม ค่าพารามิเตอร์ใหม่ให้ยีนโดยยีนในตำแหน่งใดที่จะถูก กำหนดค่าให้กำหนดจากโอกาสจากค่าความน่าจะเป็นในการกลายพันธุ์ (mutation probability) ความน่าจะเป็นในการกลายพันธุ์ (Probabilities of mutation: Pm)

โดยส่วนใหญ่ตำแหน่งและจำนวนของการ mutation จะขึ้นอยู่กับค่าการสุ่ม แต่สำหรับค่าที่ใช้ในการ mutation นั้นอาจหาค่าโดยใช้สมการที่ 4

$$After = Before + (\eta * Before) \quad (4)$$

โดยที่ After คือ ยีนหลังจากการ mutation

Before คือ ยีนก่อนการ mutation

$$\eta \text{ คือ อัตราการกลายพันธุ์ } -1 \leq \eta \leq 1$$

โดยทั่วไปค่า η กำหนดให้ค่าต่ำเพื่อไม่ให้โครโมโซม Offspring มียีนที่แตกต่างจาก Parents มากเกินไป ค่า η อาจหามาได้โดยวิธีการสุ่ม

3.6. การหยุดการเรียนรู้ (termination) เงื่อนไขในการสิ้นสุดการทำงานนั้น GA จะ สิ้นสุดการทำงานก็ต่อเมื่อได้ดำเนินการจนครบ ตามจำนวนรุ่นของประชากร (number of generation) ตามที่ผู้ใช้กำหนด หรือผู้ใช้อาจจะ กำหนดให้สิ้นสุดการทำงานเมื่อค่าคำตอบใน หลาย ๆ รอบที่ผ่านมาไม่มีการเปลี่ยนแปลง

ขั้นตอนที่ 4. การจำแนกภาพใบพืชด้วยขั้นตอนวิธีเพื่อนบ้านใกล้สุด

1. หาตัวที่ใกล้ที่สุดจากข้อมูลที่ test กับข้อมูลที่ train ซึ่งตัวที่ใกล้กับข้อมูล test มากที่สุด จะอยู่ใน class ที่ใกล้ที่สุดของข้อมูล train

2. หา k ตัวที่ใกล้ที่สุด แล้ว เอาข้อมูลที่ test ไปอยู่ใน class นั้น และถ้าเท่ากันให้ Sum Distance ของ Neighbor ของแต่ละตัว โดยจะเลือกตัวที่ใกล้ที่สุด

ขั้นตอนวิธีเพื่อนบ้านใกล้สุดจะทำการจำแนกข้อมูลที่ได้รับมาใหม่โดยอ้างอิงจากข้อมูลเดิมที่มีคุณสมบัติที่ใกล้เคียงที่สุดจำนวน k ตัวโดยจะมีขั้นตอนดังต่อไปนี้

1) D เป็นชุดข้อมูลที่ประกอบไปด้วย

$D = \{(p_0, c_0), (p_1, c_1), (p_2, c_2), \dots, (p_n, c_n)\}$ โดยที่ $pi = \{pi_1, pi_2, pi_3, \dots, pin\}$ เป็นเวกเตอร์ของข้อมูลและ ci เป็นประเภทข้อมูล pi

2) ให้ $q = \{q_1, q_2, q_3, \dots, qn\}$ เป็นเวกเตอร์ข้อมูลใหม่ที่ยังไม่ได้จำแนกประเภท

3) สำหรับทุกๆ pi ใน D สามารถคำนวณหาค่าระยะห่างระหว่าง q และ pi โดยใช้วิธียูคลิดีเนียน (Euclidian Distance, d) ดังสมการที่ 5

$$d(p_i, q) = \sqrt{\sum_{j=1}^n (p_{ij} - q_j)^2} \quad (5)$$

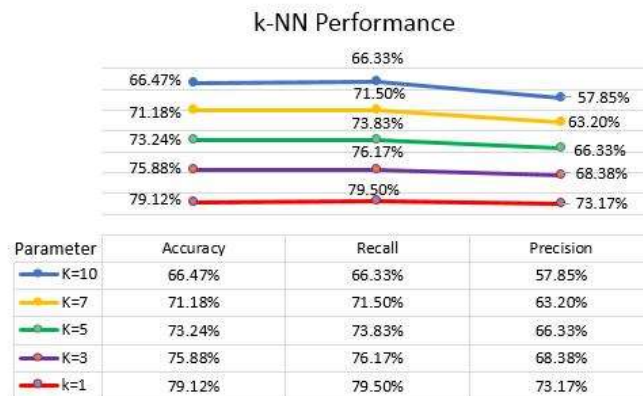
4) กำหนดประเภทของ q จากประเภทของ pi ที่มีค่าใกล้เคียงที่สุด k จำนวน โดยเลือกจากประเภทที่พบบ่อยที่สุด

ขั้นตอนที่ 5. การวัดประสิทธิภาพ

ในการวัด โดยใช้ค่าความถูกต้องของข้อมูล (accuracy) จะต้องทำการเลือกข้อมูลชุดสอน (training set) และข้อมูลชุดทดสอบ (testing set) ซึ่งในงานวิจัยนี้ใช้วิธีของ K-fold Cross-Validation ในการแบ่งกลุ่มข้อมูลออกเป็นส่วนๆ เท่าๆกัน เป็นจำนวน 10 กลุ่ม (10 fold) โดยแต่ละรอบของการคำนวณข้อมูลชุดแรกจากข้อมูลทั้ง 10 ชุด โดยข้อมูลชุดแรกจะถูกเลือกมาเพื่อเป็นข้อมูลที่ใช้ในการทดสอบ และข้อมูลชุดที่เหลือจะใช้เป็นข้อมูลสำหรับการเรียนรู้ โดยแต่ละรอบการคำนวณข้อมูลชุดหนึ่งจากข้อมูล K ชุดจะถูกเลือกออกมาเพื่อเป็นข้อมูลทดสอบ และข้อมูลอีก $K-1$ ชุดจะถูกใช้เพื่อเป็นข้อมูลสำหรับการเรียนรู้ ซึ่งข้อดีของวิธีนี้คือข้อมูลในแต่ละชุดที่ทำการแบ่งจะถูกทดสอบอย่างน้อย 1 ครั้ง และถูกเรียนรู้ทั้งหมด $K-1$ ครั้ง ในขั้นตอนสุดท้ายจะหา ค่าเฉลี่ยของค่าความถูกต้อง (accuracy) ค่าความแม่นยำ (precision) ค่าความครบถ้วน (recall) เพื่อบอกประสิทธิภาพของการจำแนกชนิดพืช

ผลการทดลอง

ผลการทดลองการจำแนกชนิดของพืชด้วยขั้นตอนวิธีเพื่อนบ้านใกล้สุดกับการเลือกตัวแทนที่เหมาะสมด้วยขั้นตอนวิธีทางพันธุกรรมโดยใช้คุณลักษณะรูปทรงและพื้นผิวของใบพืช โดยในงานวิจัยนี้ได้ทำการทดลองกับใบพืช 30 สายพันธุ์ จำนวน 340 ใบ โดยมีผลการทดลองดังภาพที่ 6



ภาพที่ 6 ผลประสิทธิภาพในการจำแนกด้วยวิธี k-NN

1) ผู้วิจัยได้ทำการทดลองโดยใช้ขั้นตอนวิธีเพื่อนบ้านใกล้สุดในการจำแนกชนิดพืชโดยไม่ใช้ตัวเลือกที่เหมาะสมจากขั้นตอนเชิงพันธุกรรมโดยใช้คุณลักษณะ 14 คุณลักษณะของใบพืช ผลการทดลองพบว่าเมื่อปรับค่าพารามิเตอร์ $K = 1$ จะได้ผลลัพธ์ที่ดีที่สุด ผลลัพธ์ความถูกต้อง (accuracy) 79.12% ค่าความครบถ้วน (recall) 79.50% และค่าความแม่นยำ (precision) 73.17%

2) ผู้วิจัยได้ทำการทดลองโดยใช้ตัวเลือกที่เหมาะสมจากขั้นตอนวิธีเชิงพันธุกรรมร่วมกับขั้นตอนวิธีเพื่อนบ้านใกล้สุดในการจำแนกชนิดพืชโดยมีผลการทดลองดังภาพที่ 7

Genetic Algorithm			Leaf	K-NN	Performance		
Pop_size	Pc	Pm	Attributes	Parameter	Accuracy	Recall	Precision
50	0.9	0.05	9	k=1	86.27%	83.06%	84.06%

ภาพที่ 7 ผลประสิทธิภาพในการจำแนกด้วยวิธี k-NN ร่วมกับ GA

ผลจากการทดลองพบว่าด้วยขบวนการขั้นตอนวิธีเชิงพันธุกรรม โดยได้ปรับค่าพารามิเตอร์ต่างๆให้เหมาะสม โดยกำหนดให้ pop size = 50 , pc = 0.9 , pm = 0.05 และจากคุณลักษณะทั้งหมด 14 คุณลักษณะของใบพืช เมื่อผ่านขบวนการเลือกตัวแทนที่เหมาะสมทำให้ได้ตัวเลือกที่เหมาะสม 9 คุณลักษณะ ดังนี้ 1. Eccentricity 2. Aspect Ratio 3. Elongation 4. Stochastic Convexity 5. Isoperimetric Factor 6. Maximal Indentation Depth 7. Average Intensity 8. Smoothness 9. Third moment ซึ่งให้ผลลัพธ์ความถูกต้องในการจำแนกชนิดของพืช (accuracy) 86.27% ค่าความครบถ้วน (recall) 83.06% และค่าความแม่นยำ (precision) 84.06%

สรุปและอภิปรายผล

ในงานวิจัยนี้ได้นำเสนอจำแนกชนิดของพืชด้วยขั้นตอนวิธีเพื่อนบ้านใกล้สุด ร่วมกับการเลือกตัวแทนที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรมโดยใช้คุณลักษณะรูปทรงและพื้นผิวของใบพืช โดยในงานวิจัยนี้ได้ทำการทดลองกับใบพืช 30 สายพันธุ์ จำนวน 340 ใบ จากผลการทดลองโดยงานวิจัยนี้ใช้ขั้นตอนวิธีเชิงพันธุกรรมในการเลือกตัวแทนที่เหมาะสม และใช้ขั้นตอนวิธีเพื่อนบ้านใกล้สุดในการจำแนกชนิดของพืชซึ่งให้ผลลัพธ์ความถูกต้อง

ในการจำแนกชนิดของพืช (accuracy) 86.27% ค่าความครบถ้วน(recall) 83.06% และค่าความแม่นยำ (precision) 84.06% เมื่อเปรียบเทียบกับการใช้ขั้นตอนวิธีเพื่อนบ้านใกล้เคียงในการจำแนกชนิดพืชเพียงอย่างเดียวซึ่งให้ผลลัพธ์ความถูกต้องในการจำแนกชนิดของพืช (accuracy) 79.12% , ค่าความครบถ้วน (recall) 79.50% และค่าความแม่นยำ (precision) 73.17% ซึ่งจากผลการทดลองพบว่า เมื่อใช้ขั้นตอนเชิงพันธุกรรมมาช่วยในการเลือกตัวแทนที่เหมาะสมทำให้ประสิทธิภาพการจำแนกชนิดพืชด้วยขั้นตอนวิธีเพื่อนบ้านใกล้เคียงมีความถูกต้องมากขึ้น มากกว่าการใช้ขั้นตอนเพื่อนบ้านใกล้เคียงเพียงอย่างเดียว

เอกสารอ้างอิง

ภาษาไทย

ดร.อาณัติ ลีมัดเดช. (2546) .การสร้างพอร์ทหุ่นด้วยวิธีคำนวณเชิงพันธุกรรม. วารสารบริหารธุรกิจ ฉบับที่ 99 (กรกฎาคม - กันยายน 2546)

ภาษาต่างประเทศ

Chou P. H., Wu M. J. and Chen K. K.(2010).Integrating Support Vector Machine and Genetic Algorithm to Implement Dynamic Wafer Quality Prediction System.Expert Systems with Applications.:4413-4424.

David E. Goldberg. (1989).Genetic Algorithm in Search Optimization and Machine Learning.Addison Wesley New York.

Daniel T. Larose.(2005).Discovering Knowledge in Data. New York : A JOHN WILEY & SONS.p.2.

Dr. R.K Bhattacharjya.(2012).Introduction To Genetic Algorithms. IIT Guwahati, pp.12.

Guo, J.Z., X.F. Wang, D.S. Huang, Z. Chi, Y.M. Cheung, J.X. Du and Y.Y. Wan. (2004).

A hypersphere method for plant leaves classification. In The Proceedings of the International Symposium on Intelligent, Multimedia, Video and SpeechProcessing. Hong Kong, China. pp. 165–168.

Huang, L. and H. Peng.(2008).Machine recognition for broad-leaved trees based onsynthetic features of leaves using probabilistic neural network. In The Proceedings of International Conference on Computer Science and Software Engineering. Wuhan, China.pp. 871–877

Huang J., Hu X. and Yang F.(2011).Support Vector Machine with Genetic Algorithm for Machinery Fault Diagnosis of High Voltage Circuit Breaker. Measurement. 44 : 1018-1027.

- Hong, F. and Z. Chi. (2003). **A two-stage approach for leaf vein extraction**. In The Proceedings of the International Conference on Neural Network & Signal Processing. Nanjing China. pp. 208–211
- Hong, F. and Z. Chi. (2006). **Combined thresholding and neural network approach for vein pattern extraction from leaf images**. The IEEE Journal of Vision Image and Signal Processing .153: 881– 892.
- Ilhan I. and Tezel G. (2013). **A Genetic Algorithm Support Vector Machine method with Parameter Optimization for Selecting the Tag SNPs**. Journal of Biomedical Informatics. 46 : 328-340.
- Jiazhi, P. and Y. He. (2008). **Recognition of plant by leaves digital image and neural network**. In The Proceedings of International Conference on Computer Science and Software Engineering. Wuhan, China. pp. 906–910.
- John H. Holland. (1975). **Adaptation in Natural and Artificial System**. University of Michigan press Ann Arbor.
- Kabir M. M., Shahjahan M. and Murase K.(2011). **A New Local Search based Hybrid Genetic Algorithm for Feature Selection**. Neurocomputing. 74 : 2914-2928.
- Padmavathi.K .(2015). **Identification of Plant Leaves disease Detection and Optimal Solution Using Genetic Algorithm**. International Journal of Advanced Research in Computer Science and Software Engineering.issn:2277 128x.
- Panagiotis, T., S.E. Papadakis and D. Manolakis. (2005). **Plant leaves classification based on morphological features and a fuzzy surface selection technique**. In The Proceedings of International Conference on Technology and Automation. Thessaloniki, Greece. pp. 365–370.
- Pedro F. B. Silva, Andre R.S. Marcal, Rubim M. Almeida da Silva (2014). **"Leaf" dataset**.
- Prof Meeta Kumar , Mrunali Kamble , Shubhada Pawar , Prajakta Patil , Neha Bonde .**Survey on Techniques for Plant Leaf Classification**. International Journal of Modern Engineering Research (IJMER) Vol.1, Issue.2, pp-538-544
- Stephen, G.W., F.S. Bao, E.Y. Xu, Y.X. Wang, Y.F. Chang and Q.L. Xiang. (2007). **A leaf recognition algorithm for plant classification using probabilistic neural network**. In The Proceedings of International Symposium on Signal Processing and Information Technology. Cairo, Egypt. , pp. 11–16.
- Tsai C. F., Eberle W. & Chu C. Y.(2013). **Genetic Algorithms in Feature and Instance Selection**. Knowledge Based Systems. 39 : 240-247

- Xiao, G., J.X. Du and X.F. Wang. (2005). **Leaf recognition based on the combination of wavelet transform and Gaussian interpolation.** In The Proceedings of International Conference on Intelligent Computer. Hefei, China. pp. 253–262.
- Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda et al. (2008). **Top 10 algorithms in data mining.** KnowlInfSyst .14:1-37.
- X. Pei-Yong, D. Xiang-Qian, and J. Bai-Ning.(2009).**A GA-based feature selection and ensemble learning for high-dimensional datasets.** in Machine Learning and Cybernetics, 2009 International Conference on. pp. 7-12.
- Yi C., Yihua L. and Haozheng R. (2012). **A Feature Selection Method Base on GA for CBIR Mammography CAD.** International Conference on the Intelligent Human Machine Systems and Cybernetics (IHMSC).
- Yun, F.L., Q.S. Zhu, Y.K. Cao and C.L. Wang. (2005). **“A leaf vein extraction method based on snakes technique.”** In The Proceedings of International Conference on Neural Networks and Brain. Beijing, China. pp. 885–888.
- Z.Michalewicz.(1972). **Genetic Algorithm + Data Structures = Evolution programs** Springer Verlag. New York.