

CP Senior Project Proposal

ข้อเสนอโครงการวิศวกรรมคอมพิวเตอร์และผลการศึกษาเบื้องต้น

เรื่อง

การจำแนกชนิดนกด้วยเสียง

Bird Song Identification

โดย

ฉันทพร พินธุพันธ์ รหัสนิสิต 5830242621

อาจารย์ที่ปรึกษา

ผศ. ดร. โปรตปราน บุญยพุกกณะ

รายงานนี้เป็นส่วนหนึ่งของวิชา 2110490 โครงการวิศวกรรมคอมพิวเตอร์พื้นฐาน

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ประจำปีการศึกษา 2561

สารบัญ

บทคัดย่อ.....	3
บทที่ 1 บทนำ.....	4
1.1. ที่มาและความสำคัญ.....	4
1.2. วัตถุประสงค์ของโครงการ.....	4
1.3. ขอบเขตของโครงการ.....	4
1.4. ขั้นตอนและแผนการดำเนินโครงการ.....	5
1.5. ประโยชน์ที่คาดว่าจะได้รับ	5
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	6
2.1. ทฤษฎีที่เกี่ยวข้อง	6
2.2. งานวิจัยที่เกี่ยวข้อง	7
บทที่ 3 วิธีการที่นำเสนอ	9
บทที่ 4 ผลการวิเคราะห์	12
บทที่ 5 บทสรุปและแนวทางในอนาคต	16
เอกสารอ้างอิง.....	17

บทคัดย่อ

เสียงร้องของนกมีประโยชน์ในการจำแนกชนิดของนกที่มีลักษณะภายนอกคล้ายคลึงกัน รวมถึงใช้ในการสำรวจความหลากหลายและความหนาแน่นของนกในแต่ละท้องถิ่น แต่เนื่องจากความคล้ายคลึงและความหลากหลายของเสียง ทำให้การจำแนกชนิดนกด้วยเสียงร้องจำเป็นต้องทำโดยผู้เชี่ยวชาญที่มีประสบการณ์ โครงการนี้จึงได้จัดทำขึ้น เพื่อศึกษาวิธีจำแนกเสียงนกแบบอัตโนมัติโดยใช้วิธีการทางคอมพิวเตอร์ และ machine learning ในการประมวลผล เพื่อให้การจำแนกชนิดนกสามารถทำได้อย่างถูกต้องแม่นยำ รวดเร็ว และทำได้โดยบุคคลทั่วไปที่มีความสนใจเกี่ยวกับนก โดยมีขอบเขตเฉพาะนกในประเทศไทยที่พบได้บ่อยในกรุงเทพมหานคร วิธีการที่นำเสนอในเบื้องต้นได้ใช้ความรู้ด้าน computer vision ในการสกัด feature ของเสียงนกจาก spectrogram และนำไปจำแนกชนิดของนกด้วยโมเดล random forest ซึ่งได้ผลลัพธ์ที่มีความแม่นยำร้อยละ 92.6 เมื่อวัดผลด้วยค่า AUC score

Abstract

Bird sounds is useful for identifying the birds with similar appearance and used for understanding the bird behavior and population trends in each location. Some of the major challenges include multiple simultaneously vocalizing birds and other sources of non-bird sound (e.g. buzzing insects). The objective of this project is to apply signal processing and machine learning to identify bird sounds in the real world with the scope of the birds in Thailand that is common in Bangkok. The methodology is to apply computer vision on the spectrogram to extract the feature from the bird sounds. Using random forest model for this multi-label classification problem achieved an AUC score of 92.6 percent.

บทที่ 1 บทนำ

1.1. ที่มาและความสำคัญ

ในประเทศไทย มีนกทั้งสิ้น 982 ชนิด (ข้อมูล ณ ปี พ.ศ. 2561) โดยแต่ละชนิดก็มีเสียงร้องที่แตกต่างกันออกไป เสียงร้องของนกจึงมีประโยชน์ในการจำแนกชนิดของนกที่มีลักษณะภายนอกคล้ายคลึงกัน และทำให้ทราบว่ามีนกอะไร อยู่ที่ใกล้ - ไกลแค่ไหน แม้จะไม่เห็นตัวหรือเห็นตัวไม่ชัดเจน ซึ่งมีประโยชน์ในการศึกษาความหลากหลายและความหนาแน่นของนกในแต่ละท้องถิ่น แต่เนื่องจากความคล้ายคลึงและความหลากหลายของเสียง ทำให้การจำแนกชนิดนกด้วยเสียงร้องจำเป็นต้องทำโดยผู้เชี่ยวชาญที่มีประสบการณ์

โครงการนี้จึงได้จัดทำขึ้น เพื่อศึกษาวิธีจำแนกเสียงนกแบบอัตโนมัติด้วยวิธีการทางคอมพิวเตอร์ โดยใช้เทคนิคต่าง ๆ เช่น machine learning, signal processing, computer vision เป็นต้น เพื่อให้การจำแนกชนิดนกสามารถทำได้อย่างถูกต้องแม่นยำ และรวดเร็ว เพื่อใช้เป็นตัวช่วยในการศึกษาเกี่ยวกับนกสำหรับผู้สนใจ รวมถึงสามารถต่อยอดโดยการนำไปใช้สำรวจประชากรนก ให้สามารถทำได้โดยทุกคนในทุกพื้นที่ โดยไม่จำเป็นต้องใช้ผู้เชี่ยวชาญเฉพาะด้าน เพื่อการดูแล ปกป้อง และอนุรักษ์นกในประเทศไทยเป็นลำดับต่อไป

1.2. วัตถุประสงค์ของโครงการ

เพื่อพัฒนาโปรแกรมสำหรับจำแนกชนิดนกจากเสียงร้อง โดยใช้วิธีการทางคอมพิวเตอร์ และ machine learning ในการประมวลผล เพื่อนำไปใช้ประโยชน์ในการศึกษาเกี่ยวกับนกในประเทศไทย และการสำรวจประชากรนกในแต่ละท้องถิ่น

1.3. ขอบเขตของโครงการ

โปรแกรมที่พัฒนาสามารถจำแนกเสียงนกในประเทศไทยได้เพียงบางชนิดเท่านั้น โดยมีความตั้งใจจะพัฒนาให้รองรับเสียงนกมากกว่า 40 ชนิดขึ้นไป แต่ในเบื้องต้นนี้ ผู้จัดทำโครงการได้ทำการทดสอบบนเสียงนกจำนวน 11 ชนิด ซึ่งเป็นนกที่พบได้บ่อยในกรุงเทพฯ และมีไฟล์เสียงตัวอย่างเป็นจำนวนมากพอสมควร ได้แก่

1. นกเขาใหญ่ (Spotted Dove; *Spilopelia chinensis*)
2. นกเขาขาว (Zebra Dove; *Geopelia striata*)
3. นกเอี้ยงสาลิ้ง (Common Myna; *Acridotheres tristis*)
4. นกกระเต็นแต้ว (Red-wattled Lapwing; *Vanellus indicus*)

5. นกกระจอกบ้าน (Eurasian Tree Sparrow; *Passer montanus*)
6. นกกาเหว่า (Asian Koel; *Eudynamys scolopaceus*)
7. นกกินปลีอกเหลือง (Olive-backed Sunbird; *Cinnyris jugularis*)
8. นกตีทอง (Coppersmith Barbet; *Psilopogon haemacephalus*)
9. นกยางเปีย (Little Egret; *Egretta garzetta*)
10. นกอีกา (Large-billed Crow; *Corvus macrorhynchos*)
11. นกอีวาบตักแต่น (Plaintive Cuckoo; *Cacomantis merulinus*)

1.4. ขั้นตอนและแผนการดำเนินโครงการ

1. รวบรวมข้อมูลเสียงนกที่พบได้บ่อยในกรุงเทพฯ เพื่อใช้ในการสร้างและทดสอบโมเดล
2. ศึกษางานวิจัยที่เกี่ยวข้อง เพื่อหาแนวทางและตัวอย่างขอบเขตของโครงการ
3. สร้างโมเดลโดยอ้างอิงจากวิธีการของงานวิจัยตัวอย่าง
4. ทดสอบโมเดลบนข้อมูลเดียวกันกับงานวิจัยตัวอย่าง เพื่อเปรียบเทียบประสิทธิภาพของโมเดล
5. นำโมเดลมาประยุกต์ใช้บนข้อมูลเสียงนกที่รวบรวมไว้ และทดสอบประสิทธิภาพของโมเดลบนข้อมูลใหม่เทียบกับข้อมูลจากงานวิจัย
6. พัฒนาโมเดลเพิ่มเติมเพื่อให้มีประสิทธิภาพดีขึ้น ถูกต้องแม่นยำมากขึ้น และรวดเร็วขึ้น
7. พัฒนาโปรแกรมในส่วนของฟังก์ชันการใช้งาน ให้สามารถใช้งานได้ง่าย และสวยงาม

1.5. ประโยชน์ที่คาดว่าจะได้รับ

สามารถนำโปรแกรมที่ได้ไปใช้จำแนกชนิดนกจากเสียงร้องตามธรรมชาติได้จริง และผลลัพธ์ที่ได้มีความแม่นยำมากกว่าร้อยละ 90 ขึ้นไป

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1. ทฤษฎีที่เกี่ยวข้อง

โครงการนี้ได้นำความรู้จากทฤษฎีต่าง ๆ ในด้าน machine learning, signal processing และ computer vision มาประยุกต์ใช้ในการทำโครงการ ดังต่อไปนี้

Spectrogram

spectrogram เป็นภาพแสดงย่านความถี่ของเสียงหรือสัญญาณอื่น ๆ ที่มีการเปลี่ยนแปลงตามเวลา เพื่อใช้ในการวิเคราะห์คลื่นเสียงประเภทต่าง ๆ โดยภาพ spectrogram จะแสดงในระนาบ 2 แกน ได้แก่ แกนความถี่ และแกนเวลา โดยแสดงค่าในแต่ละพิกัดด้วยความเข้มของสี

Short-time Fourier transform

short-time Fourier transform หรือ STFT เป็นการแปลงที่มีความสัมพันธ์กับ Fourier transform ใช้ในการหาความถี่และเฟสของช่วงใดช่วงหนึ่งของสัญญาณที่มีการเปลี่ยนแปลงไปตามเวลา โดยจะแปลงคลื่นเสียงให้อยู่ในรูปแบบ spectrogram

ในการแปลงแบบเวลาต่อเนื่อง ฟังก์ชันที่จะทำการแปลงจะถูกคูณด้วยฟังก์ชันหน้าต่าง (window function) ซึ่งเป็นฟังก์ชันที่มีค่าไม่เป็นศูนย์ในช่วงเวลาสั้น ๆ เท่านั้น การหา Fourier transform (1 มิติ) ของผลคูณนี้ ซึ่งเสมือนการเลื่อนหน้าต่างไปตามแกนเวลา จะได้ผลลัพธ์เป็นสัญญาณ 2 มิติ แสดงในรูปแบบคณิตศาสตร์ดังต่อไปนี้

$$\text{STFT}\{x(t)\} = X(\tau, \omega) = \int_{-\infty}^{\infty} x(t)w(t - \tau)e^{-j\omega t} dt$$

โดยที่ $w(t)$ เป็นฟังก์ชันหน้าต่าง ซึ่งโดยทั่วไปนิยมใช้หน้าต่างฮานน์ (Hann window) หรือหน้าต่างเกาส์ (Gaussian window) ซึ่งมีจุดกึ่งกลางที่จุดศูนย์, $x(t)$ เป็นฟังก์ชันของสัญญาณที่จะทำการแปลงหรือคลื่นเสียง และ $X(\tau, \omega)$ เป็นผลการแปลงของ $x(t)w(t-\tau)$ ซึ่งแสดงให้เห็นถึงขนาดและเฟสของสัญญาณที่เวลาและความถี่ต่าง ๆ

Random forest

Decision tree เป็นโมเดลที่ใช้ในการทำ classification ที่ใช้โครงสร้างข้อมูลรูปแบบ tree โดย inner node ใช้แสดงตัวแปรของข้อมูล, branch ใช้แสดงทางเลือกที่เกิดจากการเปรียบเทียบค่าของตัวแปร และ leaf node ใช้แสดงคลาสของข้อมูล การทำนายคลาสของข้อมูล จะเริ่มพิจารณาจาก root node ก่อน แล้วไล่ตาม branch ไปสู่ node ถัดไปเรื่อยๆ จนถึง leaf node ซึ่งจะแสดงผลการทำนาย หรือบอกคลาสของข้อมูลนำเข้า

Random forest เป็นโมเดลที่ทำการสุ่มเลือกชุดข้อมูล training data และ feature ออกมาหลาย ๆ ชุด และนำแต่ละชุดไปสร้าง Decision tree โดย Decision tree แต่ละต้นก็จะให้คำตอบออกมา แล้วโมเดลจะนำคำตอบของแต่ละต้นมารวมกัน เพื่อหาคำตอบสุดท้ายที่เหมาะสมที่สุด

AUC

AUC มาจาก Area under the curve เป็นค่าแสดงพื้นที่ใต้กราฟของ ROC (Receiver Operator Characteristic) curve ซึ่งเป็นกราฟความสัมพันธ์ระหว่าง true positive rate และ false positive rate ในการทำ classification การเปรียบเทียบประสิทธิภาพของโมเดลสำหรับ classification นิยมเปรียบเทียบกับพื้นที่ใต้โค้งหรือ AUC โดยค่า AUC ที่มากกว่าจะแสดงถึงประสิทธิภาพที่สูงกว่า

K-Fold Cross Validation

K-Fold Cross Validation เป็นวิธีการทดสอบประสิทธิภาพของโมเดล โดยทำการแบ่งข้อมูล training data ออกเป็น k ส่วน ให้แต่ละส่วนมีจำนวนข้อมูลเท่ากัน และใช้ข้อมูล 1 ส่วนเป็นตัวแทนทดสอบประสิทธิภาพของโมเดล ทำเช่นนี้วนไปจนครบ k รอบ ตามจำนวนชุดข้อมูลที่แบ่งไว้ เพื่อให้ข้อมูลทุกส่วนมีโอกาสได้ถูกทดสอบ ทำให้มีความแม่นยำและน่าเชื่อถือยิ่งขึ้น

2.2. งานวิจัยที่เกี่ยวข้อง

โครงการนี้มีการอ้างอิงวิธีการสร้างโมเดลจาก Bird Song Classification in Field Recordings: Winning Solution for NIPS4B 2013 Competition ของ Mario Lasseck [3] โดยงานวิจัยนี้เป็นส่วนหนึ่งของการแข่งขัน Multi-label Bird Species Classification - NIPS 2013 ซึ่งเป็นการแข่งขันจำแนกเสียงนกจำนวน 87 ชนิด บนไฟล์เสียง 1,000 ไฟล์ ความยาว 0.25 ถึง 5.75 วินาที ที่อัดจากสถานที่ต่าง ๆ ในฝรั่งเศส

โดยวิธีการที่ Mario ใช้ในการสร้างโมเดล เริ่มจากการนำไฟล์เสียงมาแปลงเป็น spectrogram และมองเป็นรูปแบบภาพขาวดำ เพื่อนำไปหาส่วนประกอบที่สำคัญของภาพ จากนั้นนำไปทำ classification ด้วย random forest 87 โมเดล แต่ละโมเดลแทนนกแต่ละชนิด ใช้บอกว่ามีเสียงนกชนิดนั้นอยู่ในไฟล์เสียงที่นำมาทดสอบหรือไม่ โดยหา feature จากการคำนวณค่าความคล้ายของไฟล์เสียงที่นำมาทดสอบกับชิ้นส่วนที่สำคัญของเสียงนกแต่ละชนิด ผลของงานวิจัยนี้ได้คะแนน AUC score เท่ากับ 0.91751 และถือเป็นผู้ชนะในการแข่งขันนั้น

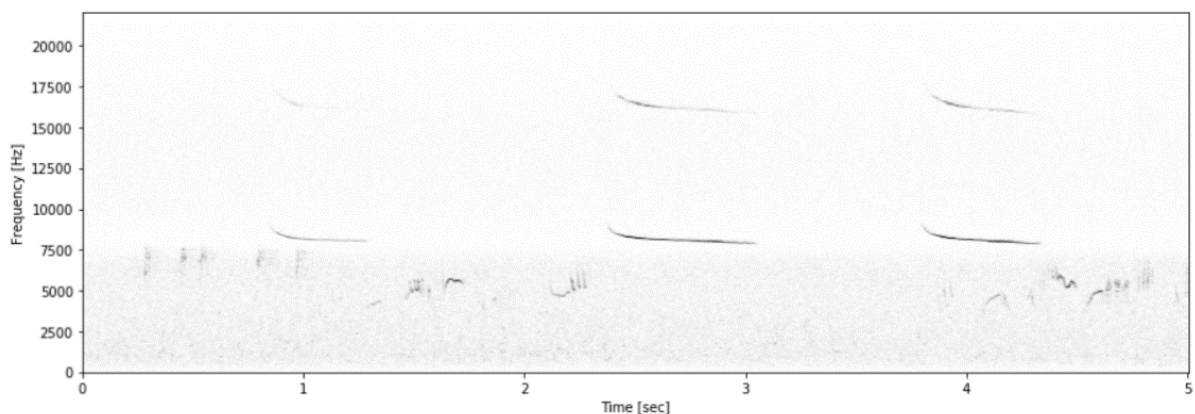
บทที่ 3 วิธีการที่นำเสนอ

โปรแกรมที่พัฒนาจะรับข้อมูลไฟล์เสียงซึ่งอยู่ในรูปแบบไฟล์ .wav แล้วนำมาประมวลผล เพื่อบอกว่าในไฟล์เสียงนั้นมีเสียงนกชนิดใดบ้าง โดยเสียงนกที่ตอบได้จะต้องอยู่ในขอบเขตของโปรแกรม ในส่วนขั้นตอนการประมวลผลของโปรแกรม ได้อ้างอิงวิธีการจากงานวิจัยของ Mario [3] โดยสามารถแบ่งขั้นตอนการประมวลผลออกเป็น 4 ขั้นตอนย่อย ดังนี้

ขั้นตอนที่ 1: Preprocessing

นำไฟล์เสียงที่จะทดสอบมาแปลงเป็น spectrogram ด้วยเทคนิค short-time Fourier transform (STFT) โดยใช้ฟังก์ชันหน้าต่างแบบ Hann window เพื่อหาความถี่ของสัญญาณเสียงที่มีการเปลี่ยนแปลงไปตามเวลา จากนั้น ทำการ normalize ค่าใน spectrogram ให้อยู่ในช่วง 0 ถึง 1 แล้วทำการคัดเลือกช่วงความถี่ให้เหลือเพียงช่วง 170 – 10,000 Hz โดยตัดช่วงความถี่ต่ำสุด 4 ช่วง และสูงสุด 24 ช่วงออก

ผลลัพธ์ในขั้นตอนนี้คือ spectrogram ในช่วงความถี่ที่เหมาะสม ซึ่งเสมือนเป็นรูปภาพขาวดำ ที่จะถูกนำไปวิเคราะห์ในขั้นตอนต่อไป

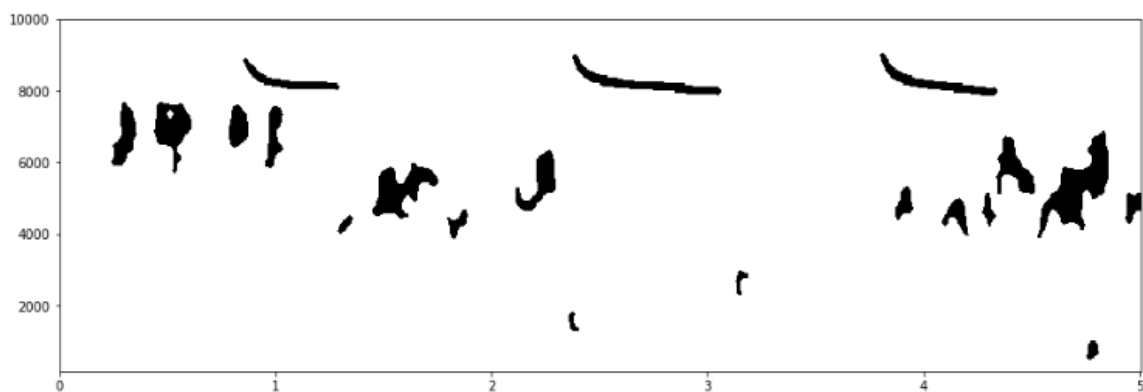


ภาพที่ 3.1 ตัวอย่างภาพ spectrogram ของไฟล์เสียงที่นำมาทดสอบ

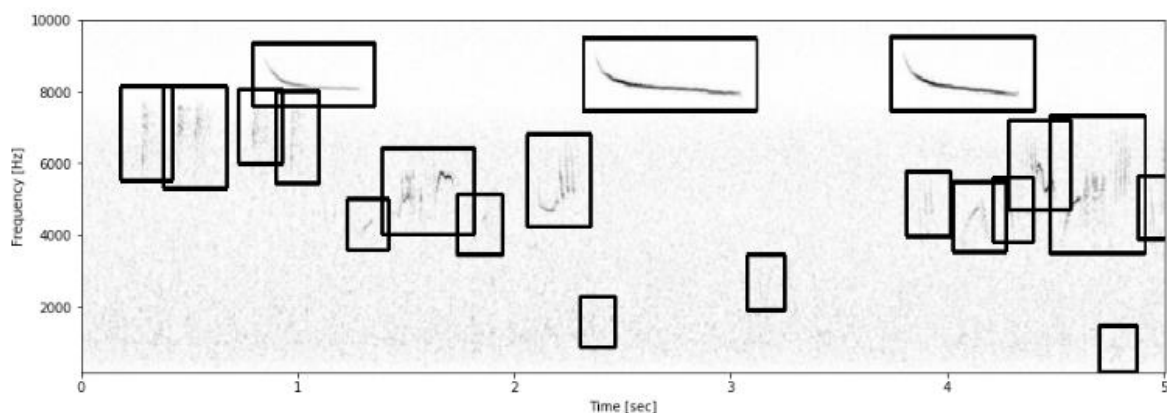
ขั้นตอนที่ 2: Segmentation

นำภาพ spectrogram ที่ได้จากขั้นตอนที่แล้วมาทำการแยกส่วนประกอบที่สำคัญของภาพ เริ่มจากการลบสัญญาณพื้นหลัง (background noise) ในภาพออกด้วยเทคนิค median clipping คือให้ pixel ที่มีค่ามากกว่า 3 เท่าของมัธยฐาน (median) ทั้งในแนวแกนความถี่และแกนเวลามีค่าเป็น 1 และ pixel ที่เหลือเป็น 0 เพื่อเก็บเฉพาะค่าที่มีความชัดเจนเท่านั้น ทำให้ภาพอยู่ในลักษณะ binary จากนั้นใช้เทคนิคการประมวลผลภาพต่าง ๆ เช่น closing, dilation, median blur ก่อนนำไปแยกส่วนประกอบโดยการหาเส้นขอบ (contour) ของแต่ละส่วนในภาพ ซึ่งจะเลือกเฉพาะชิ้นส่วนที่มีขนาดไม่เล็กจนเกินไป (ชิ้นส่วนที่เล็กมากมีโอกาสสูงที่จะเกิดจากเสียงรบกวน เช่น เสียงฝน)

ผลลัพธ์ในขั้นตอนนี้คือ รายการของภาพชิ้นส่วนสำคัญ (segment list) ของภาพ spectrogram พร้อมกับช่วงความถี่ของชิ้นส่วนนั้น



ภาพที่ 3.2 ภาพ spectrogram หลังการลบสัญญาณพื้นหลัง



ภาพที่ 3.3 ภาพแสดงชิ้นส่วนสำคัญของ spectrogram

ขั้นตอนที่ 3: Feature generation

นำเสียงของนกแต่ละชนิดมาทำขั้นตอนที่ 1 และ 2 เพื่อหาชิ้นส่วนสำคัญของเสียงของนกแยกตามชนิด จากนั้นหา feature ของแต่ละไฟล์เสียง โดยหาเปอร์เซ็นต์ความคล้ายสูงสุดของไฟล์เทียบกับแต่ละชิ้นส่วนของนกชนิดนั้น ๆ โดยจะเทียบเฉพาะช่วงความถี่ที่ใกล้เคียงกันเท่านั้น

ขั้นตอนที่ 4: Classification

เนื่องจาก 1 ไฟล์เสียง อาจประกอบด้วยนกหลายชนิด การทำ classification จึงถูกแยกออกเป็นหลายโมเดลตามจำนวนชนิดของนก กล่าวคือแต่ละโมเดลจะแทนนกชนิดหนึ่งและจะทำนายว่ามีเสียงนกชนิดนั้นอยู่ในไฟล์เสียงที่นำมาทดสอบหรือไม่ โดยใช้โมเดล random forest ในการหาค่าคะแนนความมั่นใจที่จะมีเสียงนกชนิดนั้นอยู่ในไฟล์เสียง ค่าที่ออกมาจากโมเดลจะเป็นเลขทศนิยมในช่วง 0 ถึง 1 ซึ่งถ้าค่าคะแนนมากกว่าค่า threshold หนึ่ง ๆ เราจะตอบว่ามีเสียงนกชนิดนั้นอยู่ในไฟล์เสียงที่นำมาทดสอบ

บทที่ 4 ผลการวิเคราะห์

ข้อมูลเสียงนกที่ใช้ในโครงการงานเป็นเสียงนกที่พบได้บ่อยในกรุงเทพฯ จำนวน 11 ชนิด ได้แก่

1. นกเขาใหญ่ (Spotted Dove; *Spilopelia chinensis*)
2. นกเขาขาว (Zebra Dove; *Geopelia striata*)
3. นกเอี้ยงสาธิตา (Common Myna; *Acridotheres tristis*)
4. นกกระแตแต้แว๊ด (Red-wattled Lapwing; *Vanellus indicus*)
5. นกกระจอกบ้าน (Eurasian Tree Sparrow; *Passer montanus*)
6. นกกาเหว่า (Asian Koel; *Eudynamys scolopaceus*)
7. นกกินปลีอกเหลือง (Olive-backed Sunbird; *Cinnyris jugularis*)
8. นกตีทอง (Coppersmith Barbet; *Psilopogon haemacephalus*)
9. นกยางเปีย (Little Egret; *Egretta garzetta*)
10. นกอีกา (Large-billed Crow; *Corvus macrorhynchos*)
11. นกอีวาบตักแตน (Plaintive Cuckoo; *Cacomantis merulinus*)

โดยไฟล์เสียงนกที่ใช้รวบรวมจากเว็บไซต์ www.xeno-canto.org [6] ซึ่งเป็นเว็บไซต์สำหรับแบ่งปันเสียงนกที่มีสมาชิกทั่วโลก ทำให้มีเสียงนกหลากหลายชนิดที่พบในบริเวณต่าง ๆ ทั่วโลก

ข้อมูลที่ใช้ในการสร้างและทดสอบโมเดล ได้แก่ ไฟล์เสียงนกจำนวน 120 ไฟล์ ความยาวประมาณ 4 – 40 วินาที ประกอบด้วยเสียงนก 11 ชนิด ชนิดละ 10 ไฟล์ และเสียงนกอื่น ๆ นอกเหนือจากขอบเขตของโครงการอีก 10 ไฟล์ โดยแต่ละไฟล์อาจมีเสียงนกหลายชนิดปนกันได้

ในขั้นตอนการหาส่วนประกอบสำคัญของเสียงนกแต่ละชนิด (segmentation) ใช้ไฟล์เสียงเพียง 110 ไฟล์แรกที่อยู่ในขอบเขตของโครงการ โดยตัดให้เหลือความยาวไม่เกิน 6 วินาที เพื่อให้จำนวน segment หรือจำนวน feature ไม่มากเกินไป

ผลการทดสอบโมเดล แบ่งออกเป็น 2 ขั้นตอน โดยขั้นตอนแรกเป็นผลจากการทดสอบบนข้อมูลจากงานวิจัยที่เกี่ยวข้อง และขั้นตอนที่ 2 คือผลจากการทดสอบบนข้อมูลที่ใช้ในโครงการ

ผลการทดสอบโมเดลบนข้อมูลจากงานวิจัยที่เกี่ยวข้อง

เนื่องจากงานวิจัยที่เกี่ยวข้องเป็นส่วนหนึ่งของการแข่งขัน Multi-label Bird Species Classification - NIPS 2013 ซึ่งอยู่บนเว็บไซต์ www.kaggle.com ผู้จัดทำโครงการจึงใช้ข้อมูล train data และ test data ตามการแข่งขันนั้น และนำผลการ prediction ที่ได้จากโมเดลไปส่งบนเว็บไซต์ โดยคะแนนที่ประเมินจากเว็บไซต์คำนวณจากค่า AUC score ได้เท่ากับ 0.90170 หรือคิดเป็นอันดับที่ 4 ในการแข่งขัน

ผลจากการทดสอบบนข้อมูลที่ใช้ในโครงการ

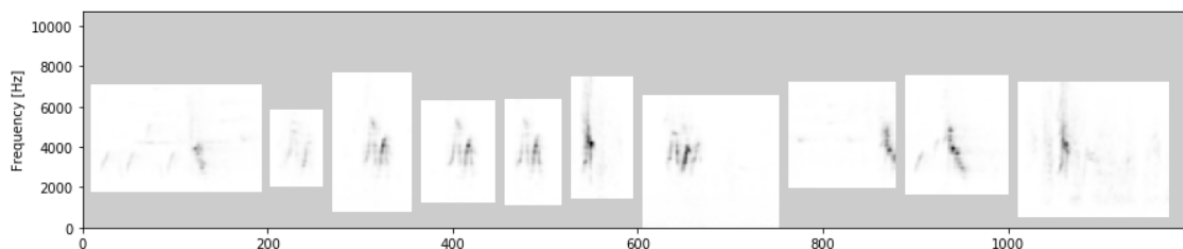
การทดสอบบนข้อมูลเสียงนกที่ใช้ในโครงการ ประเมินผลด้วยค่า AUC score โดยใช้วิธี 5-Fold Cross Validation แยกตามชนิดของนก ได้ผลลัพธ์ดังแสดงในตารางที่ 4.1

ตารางที่ 4.1 ตารางแสดงจำนวนชิ้นส่วนสำคัญ และค่า AUC score ของเสียงนกแต่ละชนิด

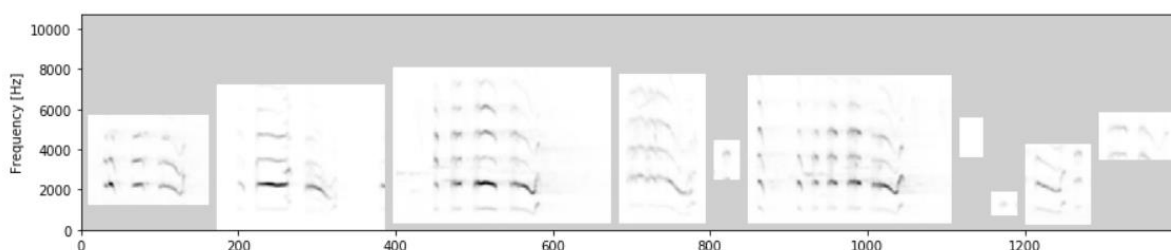
Class no	Class Name	#features	AUC score
1	นกเขาใหญ่ (<i>Spilopelia chinensis</i>)	351	0.9864
2	นกเขาชวา (<i>Geopelia striata</i>)	423	0.9477
3	นกเอี้ยงสาลิภา (<i>Acridotheres tristis</i>)	306	0.6227
4	นกกระแตแต้แว๊ด (<i>Vanellus indicus</i>)	574	0.9909
5	นกกระจอกบ้าน (<i>Passer montanus</i>)	334	0.9955
6	นกกาเหว่า (<i>Eudynamys scolopaceus</i>)	399	0.8833
7	นกกิ้งปล็อกเหลือง (<i>Cinnyris jugularis</i>)	368	0.9773
8	นกตีทอง (<i>Psilopogon haemacephalus</i>)	458	0.9591
9	นกยางเปี่ย (<i>Egretta garzetta</i>)	277	0.9045
10	นกอีกา (<i>Corvus macrorhynchos</i>)	246	0.9363
11	นกอีวาทักแตน (<i>Cacomantis merulinus</i>)	407	0.9818

จากตารางข้างต้น เมื่อนำ AUC score ของเสียงนกแต่ละชนิดมาหาค่าเฉลี่ยกัน จะได้ว่าค่า AUC score ของโมเดลนี้ เท่ากับ 0.92596 โดยจะให้เห็นว่า ค่า AUC score ของเสียงนกแต่ละชนิดมีความแตกต่างกัน โดยนกที่มีค่า AUC score สูงสุด (0.9955) คือนกกระจอกบ้าน ส่วนนกที่มีค่า AUC score ต่ำสุด (0.6227) คือนกเอี้ยงสาลิภา ซึ่งสิ่งที่ทำให้เสียงนกแต่ละชนิดมีค่า AUC score ที่แตกต่างกัน น่าจะมีผลมาจากลักษณะ

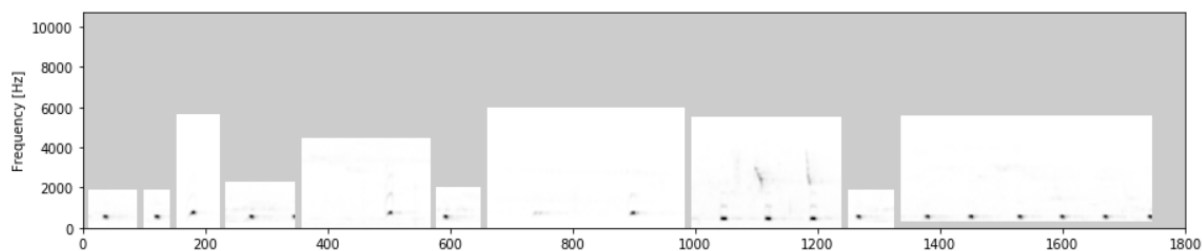
ความเฉพาะของชิ้นส่วนสำคัญของเสียงนกชนิดนั้น ๆ โดยเสียงนกที่มีชิ้นส่วนสำคัญที่มีความเฉพาะตัว ชัดเจน และมักพบซ้ำกันภายในชนิดเดียว มีแนวโน้มที่จะจำแนกได้ดีกว่าเสียงนกที่มีชิ้นส่วนสำคัญไม่ชัดเจน และแตกต่างกันภายในชนิดเดียวกัน ดังแสดงในภาพที่ 4.1 – 4.4



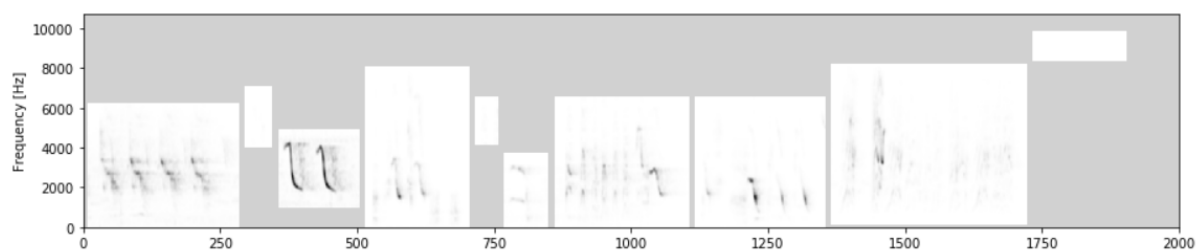
ภาพที่ 4.1 ภาพแสดงชิ้นส่วนสำคัญในการจำแนกเสียงนกกระจอกบ้าน (AUC score 0.9955)



ภาพที่ 4.2 ภาพแสดงชิ้นส่วนสำคัญในการจำแนกเสียงนกกระแตแต้แว๊ด (AUC score 0.9909)



ภาพที่ 4.3 ภาพแสดงชิ้นส่วนสำคัญในการจำแนกเสียงนกตีทอง (AUC score 0.9591)



ภาพที่ 4.4 ภาพแสดงชิ้นส่วนสำคัญในการจำแนกเสียงนกเอี้ยงสาธิต (AUC score 0.6227)

ในแง่ของ precision และ recall ของโมเดล เมื่อพิจารณาถึงประโยชน์ของโครงการในการนำไปใช้เป็นตัวช่วยในการศึกษาชนิดของนกจากเสียงร้อง โมเดลที่ได้จึงควรมีค่า recall ที่มาก เพราะถ้าหากโมเดลสามารถทำนายชนิดนกออกมาได้มาก ถึงแม้ว่าจะทำนายชนิดนกที่ไม่อยู่ในเสียงที่นำมาทดสอบออกมาบ้าง แต่ผู้ใช้งานสามารถนำชนิดนกที่โมเดลเสนอออกมา ไปค้นหาต่อเพื่อตรวจสอบความถูกต้องโมเดลเองได้ ดังนั้น ผู้จัดทำจึงเลือกค่า threshold เท่ากับ 0.11 ในการทำนายว่ามีเสียงนกชนิดนั้น ๆ อยู่ในเสียงที่นำมาทดสอบหรือไม่ ซึ่งทำให้โมเดลมีค่า recall เท่ากับ 0.7455 ขณะที่ค่า precision เท่ากับ 0.5409 ซึ่งไม่ต่ำจนเกินไป และโมเดลมีค่า accuracy เท่ากับ 0.9174 โดยค่า recall และ precision ของโมเดลสำหรับเสียงนกแต่ละชนิดมีค่าดังแสดงในตารางที่ 4.2

ตารางที่ 4.2 ตารางแสดงค่า recall และ precision ของโมเดลสำหรับเสียงนกแต่ละชนิด

Class no	Class Name	recall	precision
1	นกเขาใหญ่ (<i>Spilopelia chinensis</i>)	1.0000	0.5933
2	นกเขาชวา (<i>Geopelia striata</i>)	0.7000	0.6833
3	นกเอี้ยงสาลิภา (<i>Acridotheres tristis</i>)	0.0000	0.0000
4	นกกระแตแต้แว๊ด (<i>Vanellus indicus</i>)	0.9333	0.7000
5	นกกระจอกบ้าน (<i>Passer montanus</i>)	0.9000	0.7333
6	นกกาเหว่า (<i>Eudynamys scolopaceus</i>)	0.5333	0.5167
7	นกกิ้งปล็อกเหลือง (<i>Cinnyris jugularis</i>)	1.0000	0.6467
8	นกตีทอง (<i>Psilopogon haemacephalus</i>)	0.8000	0.3567
9	นกยางเปี่ย (<i>Egretta garzetta</i>)	0.5000	0.6000
10	นกอีกา (<i>Corvus macrorhynchos</i>)	0.9000	0.4533
11	นกอีวาทักแตน (<i>Cacomantis merulinus</i>)	0.9333	0.6667

บทที่ 5 บทสรุปและแนวทางในอนาคต

การทดสอบโมเดลในเบื้องต้นได้ผลเป็นที่น่าพอใจ เนื่องจากสามารถจำแนกชนิดนกจากเสียงร้องได้แม่นยำถึงร้อยละ 92.60 โดยสามารถจำแนกนกจำนวน 7 ชนิด (จากทั้งหมด 11 ชนิด) ได้แม่นยำมากกว่าร้อยละ 95 แต่ในขณะเดียวกัน นกบางชนิด เช่น นกเอี้ยงสาธิตา สามารถจำแนกเสียงร้องได้เพียงร้อยละ 62.27 ซึ่งผลการจำแนกของนกแต่ละชนิดจะขึ้นอยู่กับลักษณะของชิ้นส่วนสำคัญของเสียงนกชนิดนั้น ๆ

แต่ในแง่ของเวลาในการประมวลผล เวลาที่ใช้ในการจำแนกเสียงนกจะแปรผันตามความยาวของเสียงที่นำมาทดสอบ เช่น ถ้าเสียงที่นำมาทดสอบยาว 8 วินาที จะใช้เวลาประมวลผลประมาณ 10 วินาที แต่ถ้าเสียงที่นำมาทดสอบยาว 40 วินาที จะใช้เวลาประมวลผลประมาณ 43 วินาที เป็นต้น ซึ่งถือว่าใช้เวลาค่อนข้างนานหากนำไปใช้จริงในชีวิตประจำวัน

แนวทางในการพัฒนาโครงการนี้ในอนาคต อาจทำได้โดยการปรับโมเดลในการประมวลผล ให้สามารถจำแนกชนิดนกได้แม่นยำยิ่งขึ้น และทำงานได้เร็วขึ้น โดยอาจใช้เทคนิค deep learning เข้ามาช่วย รวมถึงขยายขอบเขตของโครงการให้รองรับเสียงนกมากกว่า 40 ชนิดขึ้นไป

ในส่วนฟังก์ชันของโปรแกรม อาจพัฒนาให้รองรับการรับข้อมูลเสียงและประมวลผลแบบ real-time และมีการแสดงผลเป็นชื่อนก พร้อมรูปภาพ เพื่อความสะดวกสำหรับผู้ที่ต้องการศึกษาเกี่ยวกับนก รวมถึงมีระบบเก็บข้อมูลชนิดนกที่ได้ยินในแต่ละเวลา เพื่อประโยชน์ในการนำไปสำรวจประชากรนกในท้องถิ่นต่าง ๆ เป็นลำดับต่อไป

เอกสารอ้างอิง

- [1] En.wikipedia.org. (2018). *Short-time Fourier transform*. [online] Available at:
https://en.wikipedia.org/wiki/Short-time_Fourier_transform.
- [2] Kaggle.com. (2018). *Multi-label Bird Species Classification - NIPS 2013 | Kaggle*. [online]
Available at: <https://www.kaggle.com/c/multilabel-bird-species-classification-nips2013>.
- [3] Mario Lasseck. (2013). *Bird Song Classification in Field Recordings: Winning Solution for NIPS4B 2013 Competition*. Berlin: Animal Sound Archive.
- [4] Ritchie Ng. (2018). *Evaluating a Classification Model*. [online] Available at:
<https://www.ritchieng.com/machine-learning-evaluate-classification-model/>
- [5] Scikit-learn.org. (2018). *3.1. Cross-validation: evaluating estimator performance — scikit-learn 0.20.0 documentation*. [online] Available at: https://scikit-learn.org/stable/modules/cross_validation.html.
- [6] Xeno-canto.org. (2018). *xeno-canto :: Sharing bird sounds from around the world*. [online]
Available at: <https://www.xeno-canto.org/>.