โครงการ

การเรียนการสอนเพื่อเสริมประสบการณ์

ชื่อโครงการ **บริการเพื่อการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะของเพลงจากแฟ้มข้อมูลเอ็มพี 3**

Tempo and Time Signature Detection Service from Mp3

ชื่อนิสิต **นาย พชระ เจริญกิจ 5933640323**

นาย สุรเกียรติ รอบคอบ 5933664423

ภาควิชา **คณิตศาสตร์และวิทยาการคอมพิวเตอร์**

สาขาวิชา **วิทยาการคอมพิวเตอร์**

ปีการศึกษา **2562**

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

บริการเพื่อการตรวจจับอัตราเร็วจังหวะ และเครื่องหมายกำหนดจังหวะของเพลงจากแฟ้มข้อมูลเอ็มพี 3

นาย พชระ เจริญกิจ
นาย สุรเกียรติ รอบคอบ

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิขาวิทยาการคอมพิวเตอร์ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Tempo and Time Signature Detection Service from Mp3

Patchara Charoenkij
Surakiat Robkhob

A Project Submitted in Partial Fulfillment of the Requirements
for the Degree of Bachelor of Science Program in Computer Science
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
Academic Year 2019

Copyright of Chulalongkorn University

ชื่อโครงงาน (ภาษาไทย)	บริการเพื่อการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะของเพลงจากแฟ้มข้อมูลเอ็มพี 3	
ชื่อโครงงาน (ภาษาอังกฤษ)	Tempo and Time Signature Detection Service from Mp3	
ผู้ดำเนินการ	1. นาย พชระ เจริญกิจ	เลขประจำตัวนิสิต 5933640323
	2. นาย สุรเกียรติ รอบคอบ	เลขประจำตัวนิสิต 5933664423
สาขาวิชา	วิทยาการคอมพิวเตอร์	
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. ฑิตยา หวานวารี	

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับโครงงานฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาบัณฑิต ในรายวิชา 2301499 โครงงานวิทยาศาสตร์ (Senior Project)

..... หัวหน้าภาควิชาคณิตศาสตร์

(ศาสตราจารย์ ดร. กฤษณะ เนียมมณี) และวิทยาการคอมพิวเตอร์

คณะกรรมการสอบโครงงาน

ฑิตยา หวานวารี

..... อาจารย์ที่ปรึกษาโครงงานหลัก

(ผู้ช่วยศาสตราจารย์ ดร. ฑิตยา หวานวารี)

ศศิภา พันธุวดีธร

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ศศิภา พันธุวดีธร)

รัชลิดา ลิปิกรณ์

..... กรรมการ

(รองศาสตราจารย์ ดร. รัชลิดา ลิปิกรณ์)

นาย พชระ เจริญกิจ, นาย สุรเกียรติ รอบคอบ:บริการเพื่อการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะของเพลงจากแฟ้มข้อมูลเอ็มพี 3. (Tempo and Time Signature Detection Service from Mp3) อาจารย์ที่ปรึกษาโครงงาน: ผู้ช่วยศาสตราจารย์ ดร. ฑิตยา หวานวารี, 63 หน้า.

โครงงานการวิจัยเรื่อง “บริการเพื่อการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะของเพลงจากแฟ้มข้อมูลเอ็มพี 3” มีวัตถุประสงค์ คือ พัฒนาบริการสำหรับตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะจากแฟ้มข้อมูลเอ็มพี 3 เพื่อประเมินความถูกต้องเฉลี่ยของตัวแบบ สำหรับตัวแบบที่ใช้ในการทำนายอัตราเร็วจังหวะ ทางผู้วิจัยได้ทำการทดลองปรับแต่งตัวแบบทั้งในด้านการปรับค่าอัตราการเรียนรู้ การปรับน้ำหนักของตัวแบบใหม่ตามข้อมูลนำเข้า และสำหรับตัวแบบที่ใช้ในการทำนายเครื่องหมายกำหนดจังหวะ ทางผู้วิจัยได้ทำการปรับแต่งตัวแบบในด้านการลดจำนวนชั้นของมอดูลในตัวแบบลง และการปรับน้ำหนักของตัวแบบใหม่ตามข้อมูลนำเข้า

ในด้านผลลัพธ์ของการทำนาย สำหรับการทำนายอัตราเร็วจังหวะ ตัวแบบที่เกิดจากการปรับน้ำหนักตามข้อมูลนำเข้าและค่าอัตราการเรียนรู้เป็น 0.001 นั้นให้ผลลัพธ์ได้ดีที่สุด โดยมีค่าความถูกต้องเฉลี่ยร้อยละ 73 ซึ่งมีค่าสูงกว่าการใช้ตัวแบบดั้งเดิม และสำหรับตัวแบบที่ใช้ในการทำนายเครื่องหมายกำหนดจังหวะ ตัวแบบที่ทำการปรับแต่งน้ำหนัก และตัวแบบที่มีการลดจำนวนชั้นของมอดูลลง สามารถทำนายเครื่องหมายกำหนดจังหวะได้เล็กน้อย

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์ ลายมือชื่อนิสิต .....

ลายมือชื่อนิสิต .....

สาขาวิชา วิทยาการคอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาโครงงาน .....

ปีการศึกษา 2562

5933640323, 5933664423: MAJOR MATHEMATICS AND COMPUTER SCIENCE

KEYWORDS:

Patchara Charoenkij, Surakiat Robkhob: Tempo and Time Signature Detection Service from Mp3. ADVISOR: Asst.prof. Dittaya Wanvarie, Ph.D., 63 pp.

The objective of " Tempo and Time Signature detection service from Mp3" project is to develop tempo and time signature detection service from MP3 files. To assess model accuracy, for tempo model, we adjusted model with two parameters: learning rate and model weight. For time signature model, we adjusted model with two parameters: remove module layers in model and model weight.

The tempo model with learning rate at 0.001 yields the best prediction with accuracy approximately 73%, higher than the default model. The adjusted weight model and the model with removed layers can slightly predict the time signature.

Department: Mathematics and Computer Science Student's Signature.....

Student's Signature.....

Field of Study: Computer Science Advisor's Signature .....

Academic Year: 2562

กิตติกรรมประกาศ

โครงงานบริการเพื่อการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะของเพลงจากแฟ้มข้อมูลเอ็มพี 3 สามารถสำเร็จลุล่วงไปได้ด้วยความอนุเคราะห์อย่างดียิ่งของ ผู้ช่วยศาสตราจารย์ ดร. ฑิตยา หวานวารี อาจารย์ที่ปรึกษาโครงงาน ซึ่งเสียสละเวลามาให้ความรู้ คำแนะนำ คำปรึกษาและสนับสนุนด้วยความเอาใจใส่เป็นอย่างยิ่งจนทำให้โครงงานสำเร็จลุล่วงไปได้ด้วยดี

ขอขอบพระคุณคุณ Hendrik Schreiber และคุณ Colin Raffel ที่ให้ความช่วยเหลือในการตอบข้อสงสัยที่เกี่ยวกับบทความวิจัยผ่านทางไปรษณีย์อิเลกทรอนิกส์

ขอขอบพระคุณ กรรมการสอบ ผู้ช่วยศาสตราจารย์ ศศิภา พันธุวดีธร และ รองศาสตราจารย์ ดร. รัชลิดา ลิปิกรณ์ ผู้เป็นกรรมการสอบที่ช่วยแนะแนวทางต่าง ๆ ที่เป็นประโยชน์ต่อโครงงานนี้

สุดท้ายขอขอบคุณทุกท่านที่ไม่ได้กล่าวนามไว้ข้างต้นที่ให้การสนับสนุนในด้านต่าง ๆที่คอยผลักดันให้โครงงานสำเร็จลุล่วงไปได้ด้วยดี

สารบัญ

หน้า

กิตติกรรมประกาศ..... ฉ

สารบัญตาราง..... ญ

สารบัญภาพ..... ฏ

บทที่ 1 บทนำ..... 1

1.1 ความเป็นมาและเหตุผลการวิจัย..... 1

1.2 วัตถุประสงค์ของการวิจัย..... 1

1.3 ขอบเขตการวิจัย..... 2

1.4 ขั้นตอนการวิจัย..... 2

1.5 ประโยชน์ที่ได้รับ..... 3

1.6 โครงสร้างของรายงาน..... 3

บทที่ 2 งานวิจัยและความรู้พื้นฐานที่เกี่ยวข้อง..... 4

2.1 จังหวะเคาะ อัตราเร็วจังหวะ และอัตราจังหวะ..... 4

2.2 เครื่องหมายกำหนดจังหวะ..... 5

2.3 คลังโปรแกรมที่เกี่ยวข้อง..... 6

2.4 บริการที่เกี่ยวข้อง..... 10

2.5 เมลสเปกโทรแกรม..... 11

หน้า

2.6 โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neuron Networks: CNN)..... 13

2.7 งานวิจัยที่เกี่ยวข้อง..... 14

บทที่ 3 การรวบรวมและเก็บข้อมูล..... 15

3.1 ขั้นตอนการเก็บข้อมูล..... 15

3.2 การจัดเตรียมข้อมูล..... 15

3.3 เมลสเปกโทรแกรม..... 19

บทที่ 4 ขั้นตอนการวิจัย..... 22

4.1 รูปแบบของการทำนายอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะ..... 22

4.2 โครงสร้างของโครงข่ายประสาทเทียม..... 23

4.3 การทดลอง..... 30

4.4 การวัดผลการวิจัย..... 31

4.5 การพัฒนาบริการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะ..... 31

บทที่ 5 ผลการวิจัย..... 33

5.1 ผลการดำเนินการวิจัยของการทำนายอัตราเร็วจังหวะด้วยโมเดลทั้ง 7 แบบ..... 33

5.2 ผลการดำเนินการวิจัยของการทำนายเครื่องหมายกำหนดจังหวะด้วยโมเดลทั้ง 2 แบบ..... 34

5.3 สรุปผลของการดำเนินการวิจัย..... 35

หน้า

บทที่ 6 ข้อสรุปและข้อเสนอแนะ..... 37

6.1 สรุปผลการดำเนินงาน..... 37

6.2 ปัญหาของงานวิจัยและวิธีการแก้ไข..... 37

รายการอ้างอิง..... 38

ภาคผนวก

ภาคผนวก ก คู่มือการใช้งาน..... 40

ภาคผนวก ข แบบเสนอหัวข้อโครงงาน รายวิชา 2301399 Project Proposal .....
ปีการศึกษา 2561..... 41

เอกสารอ้างอิง..... 49

ประวัติผู้เขียน..... 50

สารบัญตาราง

หน้า

ตารางที่ 3.1 แสดงคำอธิบายของตัวแปรต่าง ๆ ของเมลสเปกโทรแกรม..... 16

ตารางที่ 5.1 สรุปตัวแบบต่าง ๆ จากค่าความถูกต้องเฉลี่ย..... 33

ตารางที่ 5.2 สรุปค่าความแม่นยำ (Precision) ความครบถ้วน (Recall) และ F1-score.....
ของตัวแบบที่ 1..... 34

ตารางที่ 5.3 สรุปค่าความแม่นยำ (Precision) ความครบถ้วน (Recall) และ F1-score.....
ของตัวแบบที่ 2..... 34

ตารางที่ 5.4 สรุปค่าความแม่นยำของแต่ละตัวแบบ..... 35

สารบัญภาพ

หน้า

รูปที่ 2.1 เครื่องหมายกำหนดจังหวะ 4/4..... 5

รูปที่ 2.2 เครื่องหมายกำหนดจังหวะ 6/8..... 6

รูปที่ 2.3 ตัวเข้ารหัสค่าเดียว..... 7

รูปที่ 2.4 รายงานการจำแนกประเภท..... 8

รูปที่ 2.5 เมทริกซ์ความสับสน 8

รูปที่ 2.6 ตัวอย่างการจัดการข้อมูลโดยใช้ pandas..... 9

รูปที่ 2.7 ตัวอย่างการจัดการข้อมูลโดยใช้ pandas 9

รูปที่ 2.8 ตัวอย่างของ Spectrogram..... 12

รูปที่ 2.9 ตัวอย่างโครงข่ายประสาทเทียมแบบสังวัตนาการ [1]..... 13

รูปที่ 3.1 การแปลงแฟ้มข้อมูลเอ็มพี 3 ให้อยู่ในรูปแบบของเมลสเปกโทรแกรม..... 15

รูปที่ 3.2 ช่วงอัตราเร็วจังหวะของข้อมูลทั้งหมด..... 17

รูปที่ 3.3 รูปแบบของเครื่องหมายกำหนดจังหวะของข้อมูลทั้งหมด..... 18

รูปที่ 3.4 สเปกโทรแกรมของเพลงคำยินดี..... 19

รูปที่ 3.5 เมลสเปกโทรแกรมของเพลงคำยินดี ขนาด 256 เฟรมและไม่มีการแต่งเติม..... 19

รูปที่ 3.6 สเปกโทรแกรมของเพลงคำยินดี ขนาด 256 เฟรมและมีค่าแต่งเติมเท่ากับ 1.2..... 20

หน้า

รูปที่ 3.7 สเปกโทรแกรมของเพลงคำยินดี ขนาด 256 เฟรมและมีค่าแต่งเติมเท่ากับ 0.8..... 20

รูปที่ 3.8 สเปกโทรแกรมของเพลง คำยินดี ขนาด 512 เฟรมและไม่มีการแต่งเติม..... 21

รูปที่ 4.1 ภาพรวมของระบบ..... 22

รูปที่ 4.2 โครงสร้างโครงข่ายประสาทเทียมสังวัตนาการสำหรับการทำนายอัตราเร็วจังหวะที่มีการดัดแปลงจากงานวิจัย [1]..... 25

รูปที่ 4.3 multifilter module ในโครงสร้างโครงข่ายประสาทเทียมสังวัตนาการสำหรับการทำนายอัตราเร็วจังหวะที่มีการดัดแปลงจากงานวิจัย[1]..... 25

รูปที่ 4.4 โครงสร้างโครงข่ายประสาทเทียมสังวัตนาการสำหรับการทำนายเครื่องหมายกำหนดจังหวะที่มีการดัดแปลงจากงานวิจัย [1]..... 28

รูปที่ 4.5 multifilter module ในโครงสร้างโครงข่ายประสาทเทียมสังวัตนาการสำหรับการทำนายเครื่องหมายกำหนดจังหวะที่มีการดัดแปลงจากงานวิจัย [1]..... 28

รูปที่ 4.6 กราฟของ ELU..... 29

รูปที่ 4.7 ภาพรวมระบบบริการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะ..... 32

บทที่ 1

บทนำ

1.1 ความเป็นมาและเหตุผลการวิจัย

ดนตรีนั้นถือกำเนิดขึ้นมาอยู่กับมนุษย์มานานหลายยุคหลายสมัย ซึ่งเป็นสิ่งที่คอยช่วยให้มนุษย์เกิดความผ่อนคลายหรือช่วยให้เกิดความรื่นเริงบันเทิงใจ และมีบุคคลบางกลุ่มที่ต้องการหาโน้ตดนตรีและทำนองในบทเพลงต่าง ๆ เพื่อให้สามารถเล่นเลียนแบบต้นฉบับได้อย่างถูกต้อง

ส่วนที่สำคัญของดนตรีนั้นคือทำนองและจังหวะของดนตรี ซึ่งสองส่วนนี้เป็นส่วนหลักที่ทำให้สามารถแบ่งบทเพลงออกได้หลายประเภท หลายอารมณ์เพลง หรือหากต้องการแยกประเภทของบทเพลงต่าง ๆ ว่าสามารถจัดให้อยู่ในหมวดหมู่อะไรได้บ้างนั้น จะต้องทราบทำนองและจังหวะของดนตรีในบทเพลงนั้น ๆ ก่อน จึงจะสามารถบอกได้ว่าเพลงดังกล่าวนั้นอยู่ในหมวดหมู่อะไร

การหาทำนองและจังหวะของบทเพลงต่าง ๆ นั้นต้องอาศัยความรู้ทางด้านดนตรีหลายด้านและต้องตรวจจับจังหวะเคาะของตัวโน้ตในแต่ละห้อง และนำข้อมูลเหล่านั้นมารวมกัน หรือตรวจจับตั้งแต่จังหวะเคาะของโน้ตเพลงตัวแรกถึงโน้ตเพลงตัวสุดท้าย หลังจากนั้นจึงพิจารณาจากจังหวะเคาะที่ตรวจจับได้

ทางผู้วิจัยจึงจัดทำบริการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะนี้ขึ้นมา โดยรับข้อมูลเพลงมาในรูปแบบของแฟ้มข้อมูล เอ็มพี 3 (mp3) แล้วแปลงแฟ้มข้อมูลให้อยู่ในรูปแบบของ สเปกโทรแกรม (Spectrogram) ซึ่งเมื่อวิเคราะห์สเปกโตรแกรมแล้ว จะได้ข้อมูลในรูปแบบ JSON โดยข้อมูลที่ได้ออกมานั้นจะเป็นข้อมูลของ อัตราเร็วจังหวะ (Tempo) และ เครื่องหมายกำหนดจังหวะ (Time Signature) ของเพลงนั้น ๆ

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาบริการสำหรับตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะจากแฟ้มข้อมูลเอ็มพี 3

1.3 ขอบเขตการวิจัย

1. โครงงานนี้ศึกษารูปแบบของจังหวะเคาะ อัตราเร็วจังหวะ อัตราจังหวะ จุดเริ่มต้นของจังหวะและเครื่องหมายกำหนดจังหวะ เฉพาะดนตรีประเภท พอป ร็อกแอนด์โรล และ อาร์แอนด์บี เท่านั้น
2. ข้อมูลเข้าและข้อมูลที่ใช้แฟ้มข้อมูลเอ็มพี 3 ที่มีความยาวเต็มเพลง
3. ข้อมูลที่มีการกำกับผลลัพธ์สำหรับการฝึกสอนตัวแบบเป็นข้อมูลที่มาจาก Spotify WEB API
4. ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการสำหรับการตรวจจับจังหวะและการตรวจจับจุดเริ่มต้นของโน้ต
5. โครงงานนี้เป็นโครงงานพัฒนาบริการที่รับแฟ้มข้อมูลเสียงประเภทเอ็มพี 3 แล้วส่งออกผลลัพธ์เป็นรูปแบบ JSON ซึ่งจะระบุออกมาเป็นค่าอัตราเร็วจังหวะ (Tempo) และเครื่องหมายกำหนดจังหวะ (Time Signature) โดยมีรูปแบบดังนี้

   {“song”:”1 2 3 4 5 I love you_The Bottom Blues”,

   “tempo”: 60,

   “time signature” : “4/4”,}

1.4 ขั้นตอนการวิจัย

1. ค้นหาและศึกษาบทความรวมถึงเนื้อหาที่เกี่ยวข้องกับจังหวะทางดนตรี
2. ศึกษาเครื่องมือ โปรแกรมและเทคนิคที่ใช้ในงานวิจัย
3. รวบรวมข้อมูล เพื่อกำหนดขอบเขตของโครงงาน
4. วิเคราะห์และออกแบบวิธีการ
5. พัฒนาบริการตรวจจับอัตราเร็วจังหวะ
6. ตรวจสอบความถูกต้องและวัดประสิทธิภาพ
7. สรุปผลการดำเนินงาน
8. จัดทำเอกสาร

1.5 ประโยชน์ที่ได้รับ

ประโยชน์ที่ได้รับจากการวิจัยในครั้งนี้มีดังนี้

1. ประโยชน์ต่อผู้พัฒนาระบบ

   1.1 มีความรู้ ความเข้าใจในด้านดนตรี

   1.2 มีความรู้ ความเข้าใจในการทำบริการตรวจจับอัตราเร็วจังหวะ

   1.3 เป็นการเพิ่มพูนทักษะการเขียนโปรแกรมและการพัฒนาระบบ

   1.4 เรียนรู้การคิดวิเคราะห์วางแผนการทำงานอย่างเป็นระบบแบบแผน เพื่อให้เกิดประโยชน์สูงสุดตามทรัพยากรที่มีอยู่

   1.5 ฝึกการเรียนรู้ด้วยตนเอง การยอมรับฟังความคิดเห็นของผู้อื่น ความตรงต่อเวลา ความรับผิดชอบในหน้าที่

2. ประโยชน์ต่อผู้ใช้งานระบบ

   2.1 ผู้ใช้งานสามารถนำอัตราเร็วจังหวะที่วัดได้ไปใช้ประโยชน์ในด้านต่าง ๆ เนื่องจากผลลัพธ์ที่ได้จากการตรวจจับอัตราเร็วจังหวะ และ เครื่องหมายกำหนดจังหวะนั้นมีความถูกต้องและแม่นยำ

1.6 โครงสร้างของรายงาน

รายงานฉบับนี้ประกอบไปด้วยเนื้อหา 6 บทดังต่อไปนี้

บทที่ 1 จะกล่าวถึง ความเป็นมาและเหตุผลการวิจัย

บทที่ 2 จะกล่าวถึง งานวิจัยและความรู้พื้นฐานที่เกี่ยวข้อง

บทที่ 3 จะกล่าวถึง ขั้นตอนการรวบรวมและเก็บข้อมูล

บทที่ 4 จะกล่าวถึง ขั้นตอนการวิจัย

บทที่ 5 จะกล่าวถึง ผลการวิจัย

และบทที่ 6 จะกล่าวถึง ข้อสรุปและข้อเสนอ

บทที่ 2

งานวิจัยและความรู้พื้นฐานที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงงานวิจัยและความรู้พื้นฐานที่เกี่ยวข้องกับบริการเพื่อการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะของเพลงจากแฟ้มข้อมูลเอ็มพี 3 นี้ โดยความรู้ที่ต้องเข้าใจเป็นจำพวกแรกได้แก่ จังหวะเคาะ (Beat) อัตราเร็วจังหวะ (Tempo) อัตราจังหวะ (Time) และเครื่องหมายกำหนดจังหวะ (Time Signature) เพราะเป็นสิ่งที่คอมพิวเตอร์จะวิเคราะห์ผลออกมา จากนั้นจะเป็นความรู้ทางด้านคลังโปรแกรมต่าง ๆ ที่ทางผู้วิจัยได้มีการนำมาใช้ในการทำการวิจัยเพื่อนำไปสร้างตัวแบบ อีกทั้งยังใช้ ความรู้ทางด้านบริการออนไลน์ต่าง ๆ เพื่อนำตัวแบบที่ได้ไปสร้างเป็นบริการสำหรับตรวจจับอัตราเร็วจังหวะ และเครื่องหมายกำหนดจังหวะต่อไป

2.1 จังหวะเคาะ อัตราเร็วจังหวะ และอัตราจังหวะ

ศานติ เดชคำรณ นิยามศัพท์จังหวะเคาะ อัตราเร็วจังหวะ และอัตราจังหวะ ไว้ใน [9] ดังนี้

จังหวะเคาะ คือหน่วยนับของจังหวะที่ดำเนินไปอย่างสม่ำเสมอ เพื่อที่จะเป็นโครงสร้างให้กับจังหวะประเภทอื่น ๆ เช่น อัตราเร็วจังหวะ อัตราจังหวะ ทำนองเชิงจังหวะ จังหวะประพันธ์ เป็นต้น

อัตราเร็วจังหวะ คือค่าอัตราความสัมพันธ์ระหว่างจังหวะเคาะกับเวลา โดยมีหน่วยเป็น จังหวะเคาะต่อนาที หรือ beats per minute (bpm) หากค่าจังหวะเคาะต่อนาทีมีค่าน้อยก็ส่งผลให้เพลงนั้นเป็นเพลงช้า ในทางตรงกันข้ามถ้าหากค่าหากจังหวะเคาะต่อนาทีมีค่ามากก็ส่งผลให้เพลงเป็นเป็นเพลงเร็ว

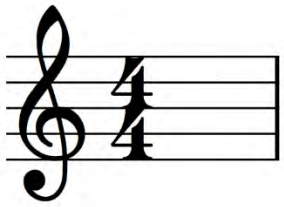
อัตราจังหวะ คือแบบแผนของจังหวะเคาะ โดยอัตราจังหวะแบ่งเป็น 3 ประเภท ได้แก่ อัตราจังหวะธรรมดา อัตราจังหวะผสม อัตราจังหวะซับซ้อน โดยอัตราจังหวะธรรมดาและอัตราจังหวะผสมสามารถพบได้ในเพลงปกติทั่วไป อีกทั้งในเพลงที่นำมาใช้ในการตรวจจับอัตราเร็วของเสียงเพลงนั้นจะเป็นเพลงที่มีสัดส่วนอัตราจังหวะธรรมดา ตามมาด้วยอัตราจังหวะผสม และสุดท้ายด้วยอัตราจังหวะซับซ้อน

โดยอัตราจังหวะธรรมดา[8] เป็นอัตราจังหวะที่ต้องไม่เป็นโน้ตประจุด แต่จะเป็นโน้ตใดก็ได้ ซึ่งอัตราจังหวะธรรมดาที่พบมากได้แก่ 2/2 3/4 4/4 เป็นต้น อัตราจังหวะผสม เป็นอัตราจังหวะที่มีโน้ตประจุดผสมอยู่ ซึ่งจะมีจังหวะตั้งแต่ 6 จังหวะ 12 จังหวะ 18 จังหวะ เป็นต้น ซึ่งอัตราจังหวะผสมที่พบมากได้แก่ 6/8 12/8 6/4

เป็นต้น และอัตราจังหวะซ้อน เป็นอัตราจังหวะที่พบไม่มากนัก โดยจะมีทั้งโน้ตธรรมดาและโน้ตประจุดปะปนกันไปโดยไม่มีหลักที่แน่นอน ตัวอย่างเช่น 7(2+2+3)/8 เป็นต้น

2.2 เครื่องหมายกำหนดจังหวะ

สำเร็จ คำโมง นิยามเครื่องหมายกำหนดจังหวะไว้ใน [6] ดังนี้

เครื่องหมายกำหนดจังหวะ (Time Signature) คือ เครื่องหมายที่เป็นตัวกำหนดจังหวะเคาะในแต่ละห้อง โดยเครื่องหมายชนิดนี้บังคับให้ห้องเพลงแต่ละห้องมีจำนวนนับจังหวะเคาะเท่ากันอย่างตายตัว ลักษณะของเครื่องหมายจะเป็นตัวเลข 2 ตัวเขียนซ้อนกัน คือให้ตัวหนึ่งวางอยู่ข้างบนอีกตัวหนึ่งโดยที่ไม่มีขีดแบ่งดังรูปที่ 1 หรือถ้าเป็นการเขียนบรรยายก็จะใช้เครื่องหมาย / คั่น เช่น 4/4, 6/8

รูปที่ 2.1 เครื่องหมายกำหนดจังหวะ 4/4

การอ่านเครื่องหมายกำหนดจังหวะแบบอัตราจังหวะธรรมดานั้น เลขตัวบนบอกจำนวนจังหวะเคาะได้ทันที และเลขตัวล่างก็บอกถึงรูปร่างของตัวโน้ตที่ใช้นับเป็น 1 จังหวะเคาะได้ทันทีโดยเลขตัวเลขสื่อความหมายดังนี้

เลข 1 ให้นับโน้ตตัวกลม (𝅝) จังหวะเคาะ

เลข 2 ให้นับโน้ตขาว (𝅗𝅥) เป็น 1 จังหวะเคาะ

เลข 4 ให้นับโน้ตตัวดำ (♩) เป็น 1 จังหวะเคาะ

เลข 8 ให้นับโน้ตตัวเขบ็ตชั้นเดียว (♪) เป็น 1 จังหวะเคาะ

เลข 16 ให้นับโน้ตตัวเขบ็ตสองชั้น (𝅘𝅥𝅯) เป็น 1 จังหวะเคาะ

และเครื่องหมายกำหนดจังหวะแบบอัตราจังหวะผสม ต้องตีความหมายของตัวเลขทั้งบนและล่างก่อนถึงจึงจะนับจำนวนจังหวะเคาะได้ โดย

เอา 3 หาร เลขตัวบน ผลลัพธ์จะเป็นจำนวนจังหวะเคาะต่อ 1 ห้อง

เอา 2 หาร เลขตัวล่าง ผลลัพธ์จะเป็นตัวโน้ตรูปร่างใดประจุด เป็น 1 จังหวะเคาะ

ตัวอย่างเช่น เครื่องหมายกำหนดจังหวะ 6/8 ดังรูปที่ 2

รูปที่ 2.2 เครื่องหมายกำหนดจังหวะ 6/8

นำ 6 ÷ 3 = 2 จะได้ว่า 1 ห้องมี 2 จังหวะ
นำ 8 ÷ 2 = 4 จะได้ว่านับเอาโน้ตตัวดำประจุด (♩.) เป็น 1 จังหวะเคาะ

2.3 คลังโปรแกรมที่เกี่ยวข้อง

ในส่วนนี้จะอธิบายเกี่ยวกับคลังโปรแกรมต่าง ๆ โดยคลังโปรแกรม LibROSA จะใช้จัดการคลื่นเสียงต่าง ๆ ในด้านการจัดการข้อมูล ได้มีการใช้คลังโปรแกรม Pandas ในการจัดการและเมื่อมีการคำนวณต่าง ๆ ได้มีการใช้คลังโปรแกรม NumPy เข้ามาสนับสนุน ส่วนด้านการจัดการการเรียนรู้ของคอมพิวเตอร์จะใช้คลังโปรแกรม TensorFlow โดยมี คลังโปรแกรม Scikit-learn คอยช่วยด้านการวิเคราะห์ผลลัพธ์และช่วย TensorFlow ดำเนินงาน

ในด้านการพัฒนาบริการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะจะใช้คลังโปรแกรม Flask และใช้ Docker เพื่อสร้างสภาพแวดล้อมจำลองให้เหมาะสมกับการดำเนินงานและนำบริการของ AWS ในด้านการให้บริการออนไลน์

LibROSA

LibROSA [5] เป็นคลังโปรแกรมที่มีขั้นตอนวิธีช่วยในการจัดการการวิเคราะห์ข้อมูลของเพลงและเสียง โดยบริการตรวจจับอัตราเร็วจังหวะนี้จะอาศัยการคัดแยกคุณลักษณะของเสียง การแปลงคลื่นเสียงให้กลายเป็นเมลสเปกโทรแกรม และใช้การตรวจจับจุดเริ่มต้นของเสียง การตรวจจับจังหวะเคาะและอัตราเร็วจังหวะ เพื่อใช้ในการเรียนรู้และการวิจัยของผู้วิจัย

NumPy

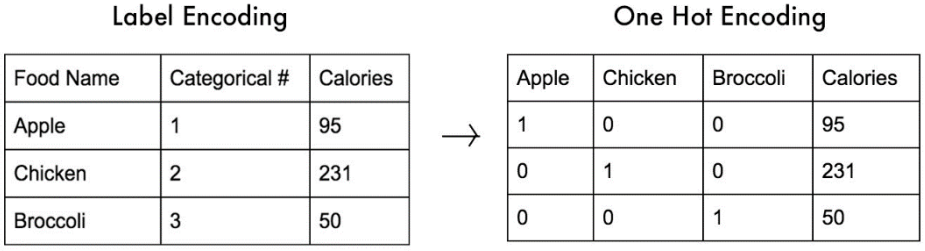
NumPy (Numerical Python) เป็นส่วนเสริมของภาษาไพธอนที่ช่วยในการคำนวณทางคณิตศาสตร์ โดยจะทำงานในรูปแบบข้อมูลแถวลำดับ (array) ซึ่งในงานวิจัยนี้จะใช้ช่วยในด้านการคำนวณตัวเลขต่าง ๆ อีกทั้งยังช่วยในการสร้างรูปแบบการสุ่มชุดข้อมูลให้เกิดข้อมูลที่มีความหลากหลายมากยิ่งขึ้น

TensorFlow

TensorFlow เป็นคลังโปรแกรมซอฟต์แวร์แบบเปิด (open-source software) ใช้สำหรับการคำนวณเชิงตัวเลขโดยอาศัยกราฟการไหลของข้อมูล (data-flow graphs) ซึ่งทางผู้วิจัยได้นำมาใช้สำหรับการเรียนรู้ของคอมพิวเตอร์

Scikit-learn

Scikit-learn เป็นคลังโปรแกรมของภาษาไพธอนที่เอื้ออำนวยแก่อัลกอริทึมประเภทการเรียนรู้แบบมีผู้สอน (supervised learning) และแบบไม่มีผู้สอน (unsupervised learning) อีกทั้งยังมีส่วนเสริมที่ช่วยในการทำงานด้านต่าง ๆ ที่นำมาใช้ร่วมกับงานวิจัยนี้ ได้แก่ การสร้างรายงานการจำแนกประเภท (classification report) การสร้างและวาดเมทริกซ์ความสับสน (confusion matrix) ตัวเข้ารหัสค่าเดียว (One Hot Encoder)



Food Name	Categorical #	Calories
Apple	1	95
Chicken	2	231
Broccoli	3	50

→

Apple	Chicken	Broccoli	Calories
1	0	0	95
0	1	0	231
0	0	1	50

รูปที่ 2.3 ตัวเข้ารหัสค่าเดียว

ที่มา: https://medium.com/@michaeldelsole/what-is-one-hot-encoding-and-how-to-do-it-f0ae272f1179

	precision	recall	f1-score	support
4-4	0.95	0.97	0.96	37
6-8	0.50	0.33	0.40	3
accuracy			0.93	40

รูปที่ 2.4 รายงานการจำแนกประเภท

Precision คือค่าความแม่นยำ สามารถคำนวณได้จากสูตร $Precision = \frac{tp}{tp+fp}$

Recall คือค่าความครบถ้วน สามารถคำนวณได้จากสูตร $Recall = \frac{tp}{tp+fn}$

โดยที่ tp (true positive) คือทำนายว่าจริง และเป็นข้อมูลจริง

tn (true negative) คือทำนายว่าเท็จ และเป็นข้อมูลเท็จ

fp (false positive) คือทำนายว่าจริง แต่ข้อมูลเท็จ

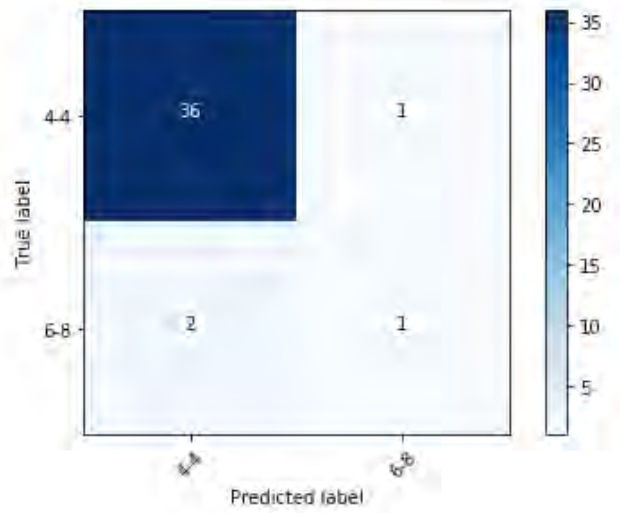
fn (false negative) คือทำนายว่าเท็จ แต่เป็นข้อมูลจริง

F1-score คือค่าเฉลี่ยของ Precision และ Recall สามารถคำนวณได้จากสูตร

$$\mathrm{F1} = 2(\frac{precision}{precision+recall})$$

Support คือจำนวนของข้อมูล

Accuracy คือค่าความแม่นยำโดยรวม สามารถคำนวณได้จากสูตร $Accuracy = \frac{tp+tn}{tp+tn+fp+fn}$

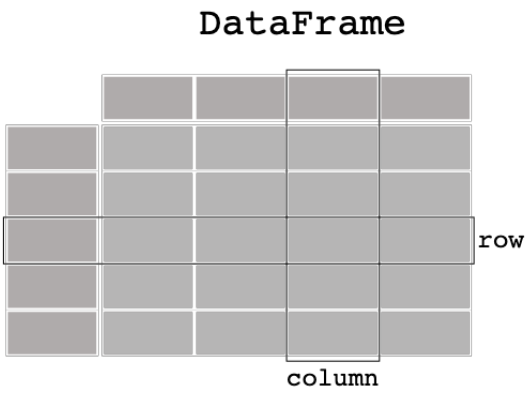


รูปที่ 2.5 เมทริกซ์ความสับสน

Pandas

Pandas คือคลังโปรแกรมของภาษาไพธอน ที่มีความสามารถในการจัดเตรียมข้อมูลไว้เพื่อสำหรับการทำงานประเภท การสร้างภาพข้อมูลแบบนามธรรม (Data Visualization) และ ใช้สำหรับการเตรียมข้อมูลในงานด้านการสร้างตัวแบบต่าง ๆ โดย Pandas นั้นถูกสร้างขึ้นบนพื้นฐานของคลังโปรแกรม NumPy

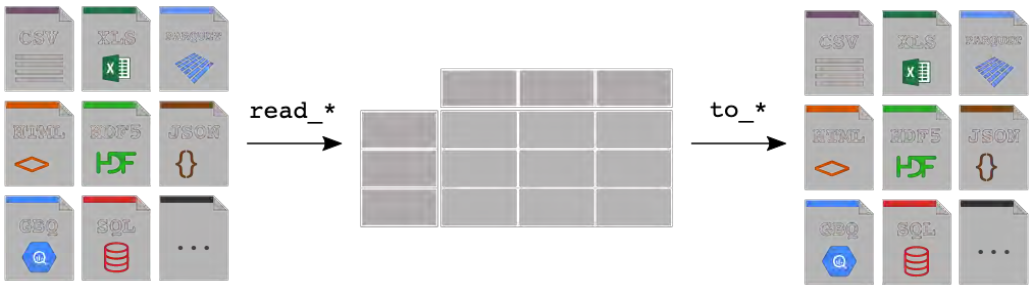
Pandas นั้นจะเข้ามาช่วยในการจัดการข้อมูลต่าง ๆ แล้วนำข้อมูลมาทำเป็นโครงข้อมูล (dataframe)



รูปที่ 2.6 ตัวอย่างการจัดการข้อมูลโดยใช้ pandas

ที่มา: https://pandas.pydata.org/docs/_images/01_table_dataframe.svg

ซึ่ง Pandas นั้นสามารถรองรับไฟล์ได้หลาย ๆ ประเภทเช่น csv, excel, sql, json และอื่น ๆ โดยการนำเข้าข้อมูลจากแหล่งข้อมูลแต่ละแหล่งจะมีคำสั่งที่เอาไว้เรียกเปิดไฟล์ประเภทต่าง ๆ นั้นคือคำสั่ง read_ * ในทำนองเดียวกันคำสั่ง to_ * จะเป็นคำสั่งที่ใช้ในการจัดเก็บข้อมูล โดยที่ * แทนด้วยประเภทของไฟล์



รูปที่ 2.7 ตัวอย่างการจัดการข้อมูลโดยใช้ pandas

ที่มา: https://pandas.pydata.org/docs/_images/02_io_readwrite.svg

Flask

Flask เป็นเว็บโครงสร้างแบบโครงคร่าว (web framework) ที่เขียนด้วยภาษาไพธอน โดย Flask ถูกเรียกว่าเป็นโครงสร้างแบบโครงคร่าวขนาดจิ๋ว (microframework) เพราะไม่ต้องอาศัยเครื่องหรือคลังโปรแกรมอื่น ๆ ก็สามารถใช้งานได้ ไม่จำเป็นต้องมีฐานข้อมูล แต่สามารถรองรับส่วนเสริมต่าง ๆ ที่ช่วยเพิ่มความสามารถของ Flask ได้ โดยในงานวิจัยนี้ ผู้วิจัยได้นำ Flask เข้ามาช่วยในการสร้างเอพีไอหรือส่วนประสานแอปพลิเคชัน

Docker

Docker เป็นแพลตฟอร์ม ที่มีการทำงานโดยการสร้างสภาพแวดล้อมจำลองขึ้นมา โดย Docker จะมี Docker image ที่ช่วยในการติดตั้งโปรแกรมสนับสนุนต่าง ๆ ที่ทางผู้วิจัยต้องการ หลังจากนั้นจึงมีการทำงานในส่วนของ Docker container ซึ่งเปรียบเสมือนกล่องที่รองรับ Docker image ต่าง ๆ

2.4 บริการที่เกี่ยวข้อง

Amazon Web Service (AWS)

Amazon SageMaker

Amazon SageMaker เป็นบริการที่ช่วยในการสร้าง ฝึก และติดตั้งตัวแบบการเรียนรู้ของเครื่อง (Machine Learning: ML) ซึ่งจะทำงานโดยการลดการทำงานที่ยุ่งยากในแต่ละขั้นตอนของกระบวนการการเรียนรู้ของเครื่องคอมพิวเตอร์ เพื่อช่วยให้การพัฒนาตัวแบบมีคุณภาพสูงขึ้น

โดย Amazon SageMaker นั้นจะรวบรวมเครื่องมือต่าง ๆ ที่ทางผู้วิจัยจำเป็นต้องใช้เพื่อพัฒนาตัวแบบ และเพื่อลดความซับซ้อนของการเรียนรู้ของคอมพิวเตอร์ ซึ่งเป็นผลให้มีค่าใช้จ่ายที่น้อยลงตามมาด้วย

Amazon Simple Storage Service (Amazon S3)

Amazon S3 เป็นบริการที่ใช้ในการจัดเก็บข้อมูลต่าง ๆ โดยทางผู้วิจัยได้ทำการนำข้อมูลขึ้นจัดเก็บออนไลน์

Amazon Elastic Container Registry (Amazon ECR)

Amazon ECR เป็นบริการจัดเก็บ Docker image โดยทางผู้วิจัยได้ทำการนำ Docker image ขึ้นจัดเก็บออนไลน์เพื่อนำไปใช้กับ Amazon Elastic Beanstalk ต่อไป

Amazon Elastic Compute Cloud (Amazon EC2)

Amazon EC2 เป็นบริการที่ช่วยให้ผู้ใช้บริการสามารถใช้เป็นคอมพิวเตอร์จำลองที่สามารถปรับแต่งได้ในระดับโครงสร้างพื้นฐาน (infrastructure) โดยสำหรับงานวิจัยนี้ Amazon EC2 จะถูกใช้เป็นผู้ให้บริการแม่ข่ายสำหรับ Amazon Elastic Beanstalk

Amazon Elastic Beanstalk

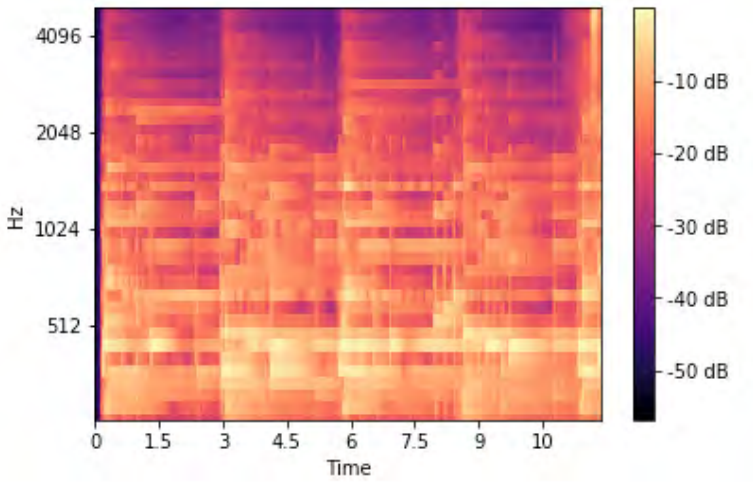
Amazon Elastic Beanstalk เป็นบริการที่ช่วยจัดการด้านการติดตั้งเว็บแอปพลิเคชัน โดยจะช่วยจัดการด้านปรับแต่งในด้านต่าง ๆ ให้อย่างอัตโนมัติ เช่น การเพิ่มหน่วยประมวลผลของ Amazon EC2 การทำ load balancer เพื่อช่วยประมวลผลเมื่อมีผู้ใช้เข้ามาจำนวนมาก สำหรับการวิจัยนี้ผู้วิจัยได้ใช้ Amazon Elastic Beanstalk เป็นผู้ช่วยในการจัดการการติดตั้งเอพีไอหรือส่วนประสานแอปพลิเคชันเพื่อให้สามารถเรียกใช้บริการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะได้

2.5 เมลสเปกโทรแกรม

2.5.1 สเปกโทรแกรม

สเปกโทรแกรม คือการแสดงภาพของสเปกตรัมความถี่ของสัญญาณที่แตกต่างกันไปตามเวลาออกมาในรูปแบบนามธรรม

โดยสเปกโทรแกรมได้มาจากการแปลงสัญญาณเสียงให้อยู่ในรูปแบบคลื่นด้วยการแปลงฟูเรียร์ (Fourier Transform) จะได้เป็นกราฟ 2 มิติ ที่ในแนวแกน y จะระบุด้วยค่าความถี่ และแนวแกน x จะระบุด้วยค่าเวลา โดยมีความเข้มของสีแสดงถึงความเข้มของความถี่ในจุดนั้น ๆ ดังรูปที่ 2.8



รูปที่ 2.8 ตัวอย่างของ Spectrogram

2.5.2 สเกลเมล (Mel Scale)

สเกลเมล เป็นหน่วยของเสียงชนิดหนึ่ง ซึ่งเกิดจากการปรับค่าความถี่ธรรมชาติให้อยู่ในระดับที่สามารถรับรู้ได้ ซึ่งจุดอ้างอิงระหว่างสเกลเมลและความถี่ธรรมชาติคือที่ 1,000 เมลจะมีค่าเท่ากับ 1,000 เฮริทซ์ โดยการแปลงความถี่ธรรมชาติให้อยู่ในสเกลเมลสามารถทำได้ดังสมการ

$$Mel = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

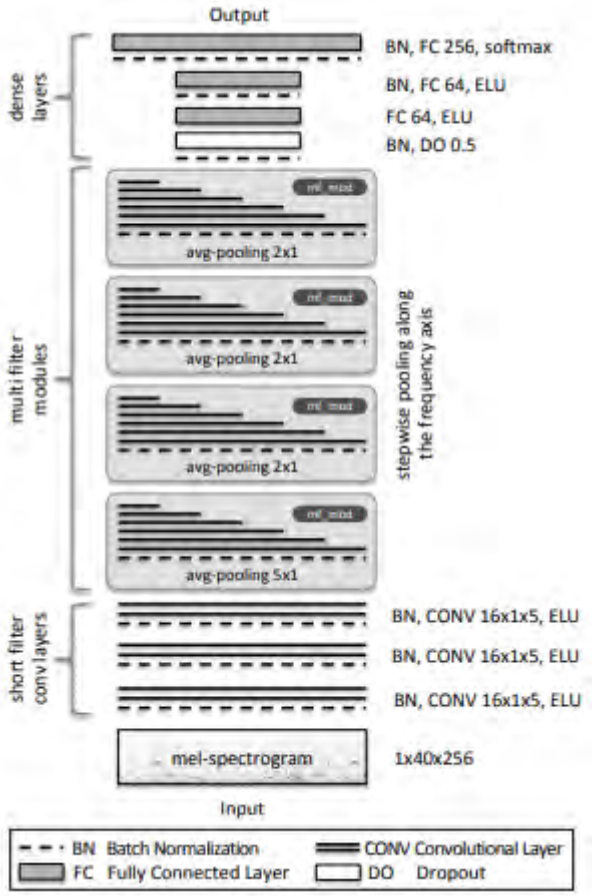
โดย f หมายถึงค่าความถี่ธรรมชาติ

ดังนั้น เมลสเปกโทรแกรม จึงหมายถึงสเปกโทรแกรมซึ่งแสดงช่วงความถี่ในสเกลเมลนั่นเอง จะเห็นได้ว่า สเกลเมลนั้นเป็นสเกลลอการิทึม เมื่อวาดสเปกโทรแกรมจะได้แกน y ซึ่งแทนความถี่ในสเกลลอการิทึมเช่นกัน

2.6 โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neuron Networks: CNN)

โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neuron Networks: CNN) เป็นโครงข่ายประสาทเทียมชนิดพิเศษ ที่ถูกนำไปใช้อย่างมีประสิทธิภาพมากในงานด้านการจดจำและจำแนกรูปภาพต่าง ๆ

CNN สามารถเป็นตัวจับลักษณะสำคัญได้โดยแปลงข้อมูลที่รับเข้ามาไปเป็นลักษณะสำคัญที่มีระดับสูงขึ้นได้ โดยข้อมูลที่รับเข้ามานั้นมาจากการผ่านชั้นต่าง ๆ ทั้งชั้นสังวัตนาการ (convolutional layer) ซึ่งจะช่วยจับลักษณะสำคัญจากข้อมูลที่ได้รับเข้ามาแล้วส่งต่อไปให้ชั้นบ่อรวม (pooling layer) ซึ่งจะทำหน้าที่ลดขนาดของข้อมูลเพื่อเอาเฉพาะข้อมูลที่มีความสำคัญเอาไว้ และปิดท้ายด้วยชั้นการเชื่อมโยงแบบเต็มรูปแบบ (fully connected layer) ซึ่งมักจะมีหลายชั้น



รูปที่ 2.9 ตัวอย่างโครงข่ายประสาทเทียมแบบสังวัตนาการ [1]

2.7 งานวิจัยที่เกี่ยวข้อง

จากงานวิจัยของ Hendrik Schreiber และ Meinard Müller [1] กล่าวไว้ว่าวิธีการทั่วไปที่จะประมาณอัตราเร็วจังหวะ (tempo) จะทำโดยการหาจุดเริ่มต้นของเสียง (onset) และ จังหวะเคาะ (beat) จากนั้นจึงนำไปประมาณหาอัตราเร็วจังหวะ แต่สำหรับงานวิจัย [1] นี้จะประมาณอัตราเร็วจังหวะโดยตรงจากสเปกโตรแกรมของเมล (mel-spectrogram) แบบขั้นตอนเดียว โดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ(convolutional neural network) ซึ่งการประมาณค่าอัตราเร็วจังหวะนี้เป็นปัญหาการจำแนกประเภทแบบหลายหมวดหมู่ (multi-class classification) โดยการสอนโครงข่ายประสาทเทียมนั้นจะใช้ชุดข้อมูลเพลงที่ครอบคลุมแนวเพลงและอัตราเร็วจังหวะที่มีความหลากหลายควบคู่ไปกับการเพิ่มข้อมูล ผลการทดลองพบว่าการประมาณค่าอัตราเร็วจังหวะสามารถเอาชนะขั้นตอนวิธีที่ดีที่สุดในตอนนั้น แต่ยังไม่สามารถเอาชนะในกรณีที่มีความคลาดเคลื่อนของอัตราเร็วจังหวะคู่แปด (tempo octave error) ที่ความคลาดเคลื่อนยินยอมที่ 4% โดยความคลาดเคลื่อนของอัตราเร็วจังหวะคู่แปดคือปัญหาของผลลัพธ์ของขั้นตอนวิธีที่ได้อัตราเร็วจังหวะออกมาเป็นเศษส่วนหรือเป็นจำนวนที่เป็นทวีคูณ

จากงานวิจัยของ Jan Schlüter และ Sebastian Böck [2] ระบุไว้ว่าจุดเริ่มต้นของเสียงคือจุดเริ่มต้นที่มีการเกิดเสียงแต่ละเสียงในเพลง โดยการตรวจจับจุดเริ่มต้นของเสียง (onset detection) สามารถนำไปใช้ประโยชน์ในงานต่าง ๆ ได้ เช่น การตรวจจับจังหวะเคาะ การประมาณค่าอัตราเร็วจังหวะ และการวิเคราะห์เพลง เป็นต้น จากสเปกโตรแกรมของเพลงจะทราบว่าจุดเริ่มต้นของเสียงเป็นการเปลี่ยนแปลงของสเปกตรัมต่อเวลา เมื่อพิจารณาจากจุดนั้นทำให้ทราบว่าการหาจุดเริ่มต้นของเสียงนั้นมีความคล้ายคลึงกับการหาขอบในรูปภาพ จึงนำไปสู่การใช้โครงข่ายประสาทเทียมแบบสังวัตนาการในการหาจุดเริ่มต้นของเสียง ผลการทดลองได้ค่าความแม่นยำ (precision) 90.5% ค่าความครบถ้วน (recall) 86.6% และค่า F-measure 88.5%

บทที่ 3

การรวบรวมและเก็บข้อมูล

ในบทนี้จะกล่าวถึง วิธีการนำเข้าแฟ้มข้อมูลประเภทเอ็มพี 3 และการเตรียมข้อมูลก่อนนำไปใช้ โดยข้อมูลที่จะนำไปให้คอมพิวเตอร์เรียนรู้ จะถูกปรับให้อยู่ในรูปแบบของเมลสเปกโทรแกรมเพื่อให้คอมพิวเตอร์สามารถวิเคราะห์อัตราเร็วจังหวะ และเครื่องหมายกำหนดจังหวะของข้อมูลที่มีได้อย่างมีประสิทธิภาพมากขึ้น

3.1 ขั้นตอนการเก็บข้อมูล

1. จัดหาแฟ้มข้อมูลประเภทเอ็มพี 3 ผ่านทางช่องทางออนไลน์ของบริษัท GMM Grammy จำกัด (มหาชน) โดยได้รับเป็นชุดข้อมูลทั้งหมด 3 ชุด ชุดละ 80 เพลง รวมทั้งหมด 240 เพลง
2. จากนั้นผู้วิจัยได้ทำการนับอัตราเร็วจังหวะและบันทึกค่าเครื่องหมายกำหนดจังหวะด้วยตัวเองเพื่อเป็นป้ายกำกับของแต่ละเพลง

3.2 การจัดเตรียมข้อมูล

1. ตรวจสอบแฟ้มข้อมูลที่มีอยู่ว่ามีข้อมูลที่ซ้ำเกิดขึ้นหรือไม่ เนื่องจากในแต่ละชุดข้อมูลเพลงที่ได้ทำการจัดหามานั้น อาจมีเพลง
2. จากข้อมูลทั้งหมดมีข้อมูลซ้ำอยู่จำนวน 42 ข้อมูล จึงทำการลบข้อมูลที่ซ้ำออก และเหลือข้อมูลทั้งสิ้น 198 ข้อมูล
3. ทำการแปลงข้อมูลจากแฟ้มข้อมูลเอ็มพี 3 ให้อยู่ในรูปแบบของเมลสเปกโทรแกรม โดยเอาแฟ้มข้อมูลมาอ่านด้วยคลังโปรแกรม LibROSA ด้วยคำสั่ง load จะได้ค่า y ที่เป็นอนุกรมเวลาของเสียงออกมา ตัวอย่างเช่น (5331937,) ซึ่งเป็นแถวลำดับ (Array) 1 มิติ และอัตราการสุ่มตัวอย่าง โดยทำการกำหนดค่าต่าง ๆ

```
def extract_tempo_features(file, window_length=1024):
    y, sr = librosa.load(file, sr=11025)
    hop_length = window_length // 2
    data = librosa.feature.melspectrogram(y=y, sr=11025, n_fft=window_length, hop_length=hop_length,
                                          power=1, n_mels=40, fmin=20, fmax=5000)
    data = np.reshape(data, (data.shape[0], data.shape[1], 1))

    return data.astype(np.float16)
```

รูปที่ 3.1 การแปลงแฟ้มข้อมูลเอ็มพี 3 ให้อยู่ในรูปแบบของเมลสเปกโทรแกรม

โดยตัวแปรต่าง ๆ จะอธิบายได้ดังตารางต่อไปนี้

ตารางที่ 3.1 แสดงคำอธิบายของตัวแปรต่าง ๆ ของเมลสเปกโทรแกรม

window_length	ค่าขนาดของเฟรมเสียง ปกติจะตั้งให้เท่ากับค่า n_ftt ซึ่งจะถูกกำหนดไว้ที่ 1024
hop_length	ค่าก้าวกระโดด เป็นค่าที่กำหนดว่าเราควรจะทำเมลสเปกโทรแกรมที่ต่อไป ณ ที่ใด ถูกกำหนดไว้ที่ 512
n_fft	ค่าของการแปลงฟูเรียร์แบบรวดเร็ว ถูกกำหนดไว้ที่ 1024
n_mels	ค่าจำนวนของเมลสเปกโทรแกรมของแต่ละเพลง กำหนดไว้ที่ 40
sr	ค่าอัตราการสุ่มตัวอย่าง กำหนดไว้ที่ 11025
fmin	ค่าความถี่ต่ำสุด กำหนดไว้ที่ 20
fmax	ค่าความถี่สูงสุด กำหนดไว้ที่ 5000

3.2.1 การแบ่งชุดข้อมูลสำหรับการวิเคราะห์อัตราเร็วจังหวะ

3.2.1.1 การแบ่งข้อมูล ข้อมูลเพลงทั้งหมด 198 เพลง แบ่งออกเป็นชุดข้อมูลสำหรับการเรียนรู้ (training set) และชุดข้อมูลสำหรับการทดสอบ (test set) ในอัตราส่วน 8 : 2 ซึ่งจะได้ชุดข้อมูลสำหรับการเรียนรู้ทั้งหมด 158 เพลง และชุดข้อมูลสำหรับการทดสอบทั้งหมด 40 เพลง

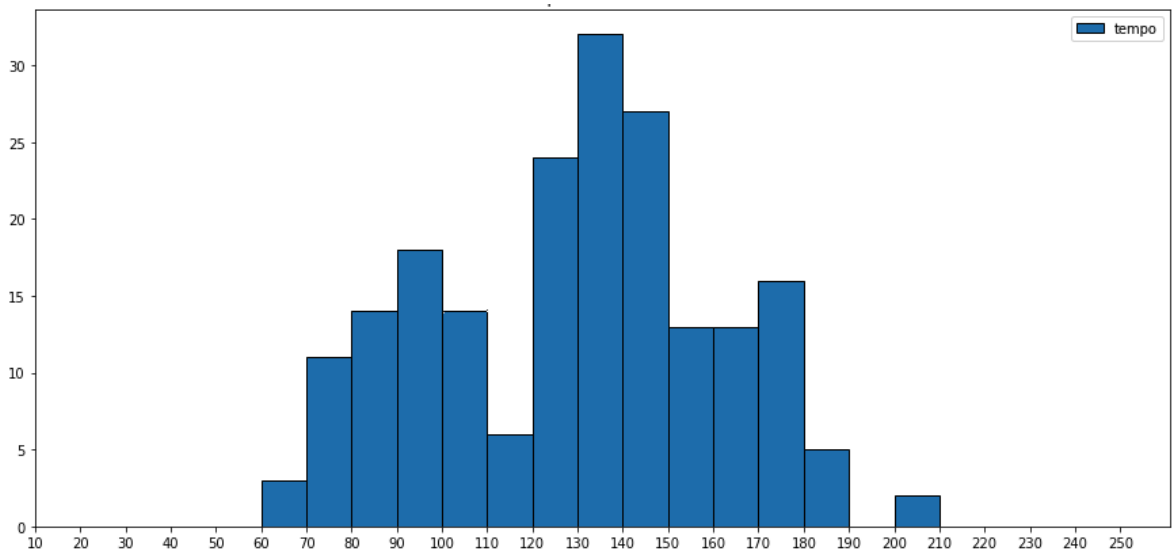
3.2.1.2 ในชุดข้อมูลสำหรับการเรียนรู้ 158 เพลงนั้น ใน 1 เพลงจะมีการตัดแบ่งด้วยวิธีการสุ่มออกมาใช้ 10 ช่วง ช่วงละ 11.9 วินาที หรือ 256 เฟรม ซึ่งมีโอกาสที่จะซ้อนทับกันได้ จากวิธีนี้จะทำให้มีชุดข้อมูลสำหรับการเรียนรู้ทั้งหมด 1580 เพลงซึ่งชุดข้อมูลนี้จะถูกแต่งเติมด้วยค่าแต่งเติมซึ่งอยู่ในช่วง 0.8 ถึง 1.2 โดยถ้าค่าแต่งเติมมีค่ามากกว่า 1 จะเป็นการบีบสเปกโทรแกรมในแนวแกน x ในทางตรงข้ามหากค่าแต่งเติมมีค่าน้อยกว่า 1 จะเป็นการยืดสเปกโทรแกรมในแนวแกน x (ดังตัวอย่างในหัวข้อ 3.3) เพื่อให้ชุดข้อมูลของอัตราเร็วจังหวะมีการกระจายมากขึ้น โดยจากวิธีการข้างต้นจะทำให้ข้อมูลเพลงที่เป็นเพลงเดียวกันมีค่าอัตราเร็วจังหวะที่แตกต่างกัน

3.2.1.3 หลังจากที่ได้ข้อมูลสำหรับการเรียนรู้ทั้งหมด 1580 เพลงแล้ว จากนั้นจะทำการสุ่มแบ่งข้อมูลออกเป็นชุดข้อมูลสำหรับการเรียนรู้ (training set) และชุดข้อมูลสำหรับการตรวจสอบ

(validate set) ในอัตราส่วน 9 : 1 ซึ่งจะได้ชุดข้อมูลสำหรับการเรียนรู้ทั้งหมด 1422 เพลง และชุดข้อมูลสำหรับการตรวจสอบทั้งหมด 158 เพลง

3.2.1.4 โดยชุดข้อมูลสำหรับการทดสอบ 40 เพลงนั้นได้จากการแบ่งข้อมูลในขั้นตอนที่ 3.2.1.1 เพื่อหลีกเลี่ยงข้อมูลของชุดทดสอบไม่ให้อยู่ในชุดข้อมูลสำหรับเรียนรู้และชุดข้อมูลสำหรับตรวจสอบ

ซึ่งจะทำกระบวนการตั้งแต่ข้อ 3.2.1.1 - 3.2.1.3 ทั้งหมด 10 รอบสุดท้ายจะได้ ชุดข้อมูลสำหรับการเรียนรู้ 10 ชุด ชุดข้อมูลสำหรับการตรวจสอบ 10 ชุด และชุดข้อมูลสำหรับการทดสอบ 10 ชุด โดยชุดข้อมูลทั้ง 10 ชุดจะนำไปใช้สำหรับตัวแบบ 1 แบบเพื่อหาค่าเฉลี่ยของชุดข้อมูลที่นำไปทดสอบ



รูปที่ 3.2 ช่วงอัตราเร็วจังหวะของข้อมูลทั้งหมด

ซึ่งรูปที่ 3.2 แสดงถึงการกระจายตัวของข้อมูลเพลง โดยที่แกน x คืออัตราเร็วจังหวะ (จังหวะเคาะต่อนาที: BPM) และแกน y คือจำนวนเพลง

3.2.2 การแบ่งชุดข้อมูลสำหรับสำหรับการวิเคราะห์เครื่องหมายกำหนดจังหวะ

3.2.2.1 การแบ่งข้อมูล จากข้อมูลเพลงทั้งหมดแบ่งออกเป็นเพลงที่มีเครื่องหมายกำหนดจังหวะเป็น 4/4 จำนวนทั้งสิ้น 191 เพลง และเพลงที่มีเครื่องหมายกำหนดจังหวะเป็น 6/8 อีกจำนวน 7 เพลง รวมข้อมูลทั้งหมด 198 เพลง

3.2.2.2 ทำการสุ่มแบ่งข้อมูลออกเป็น 2 ชุดคือชุดข้อมูลสำหรับการเรียนรู้ และชุดข้อมูลสำหรับการตรวจสอบ โดยจะแบ่งให้ ชุดข้อมูลสำหรับการเรียนรู้มีเพลงที่มีเครื่องหมายกำหนดจังหวะเป็น 6/8 อยู่จำนวน 4 เพลง และให้ชุดข้อมูลสำหรับการทดสอบเพลงที่มี

เครื่องหมายกำหนดจังหวะเป็น 6/8 อยู่จำนวน 3 เพลง ซึ่งจะทำให้ได้ชุดข้อมูลสำหรับการเรียนรู้ทั้งหมด 158 เพลงและชุดข้อมูลสำหรับการตรวจสอบ 40 เพลง

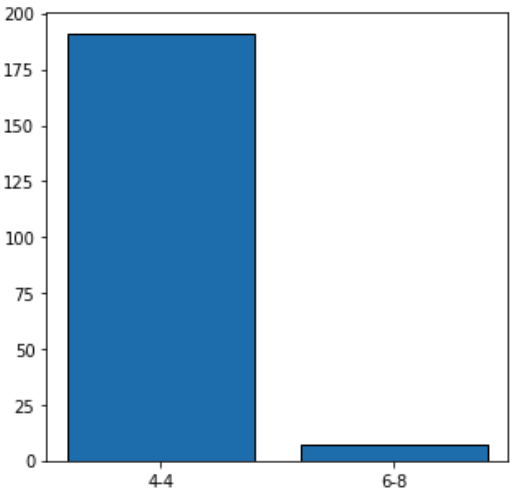
3.2.2.3 โดยในชุดข้อมูลสำหรับการเรียนรู้ 158 เพลงนั้น ใน 1 เพลงจะมีการสุ่มออกมาใช้ 10 ช่วง ช่วงละ 11.9 วินาที หรือ 256 เฟรม ซึ่งมีโอกาสที่จะซ้อนทับกันได้ จากวิธีนี้จะทำให้มีชุดข้อมูลสำหรับการเรียนรู้ทั้งหมด 1580 เพลง

3.2.2.4 หลังจากที่ได้ข้อมูลสำหรับการเรียนรู้ทั้งหมด 1580 เพลงแล้ว จากนั้นจะทำการสุ่มแบ่งข้อมูลออกเป็นชุดข้อมูลสำหรับการเรียนรู้ (training set) และชุดข้อมูลสำหรับการตรวจสอบ (validate set) ในอัตราส่วน 9 : 1 ซึ่งจะได้ชุดข้อมูลสำหรับการเรียนรู้ทั้งหมด 1422 เพลง และชุดข้อมูลสำหรับการตรวจสอบทั้งหมด 158 เพลง

3.2.1.4 โดยชุดข้อมูลสำหรับการทดสอบ 40 เพลงนั้นได้จากการแบ่งข้อมูลในขั้นตอนที่ 3.2.1.1 เพื่อหลีกเลี่ยงข้อมูลของชุดทดสอบไม่ให้อยู่ในชุดข้อมูลสำหรับเรียนรู้และชุดข้อมูลสำหรับตรวจสอบ

ซึ่งจะทำกระบวนการตั้งแต่ข้อ 3.2.2.1 - 3.2.2.4 ทั้งหมด 10 รอบสุดท้ายจะได้ ชุดข้อมูลสำหรับการเรียนรู้ 10 ชุด ชุดข้อมูลสำหรับการตรวจสอบ 10 ชุด และชุดข้อมูลสำหรับการทดสอบ 10 ชุด

โดยชุดข้อมูลทั้ง 10 ชุดจะนำไปใช้สำหรับตัวแบบ 1 แบบเพื่อหาค่าเฉลี่ยของชุดข้อมูลที่นำไปทดสอบ

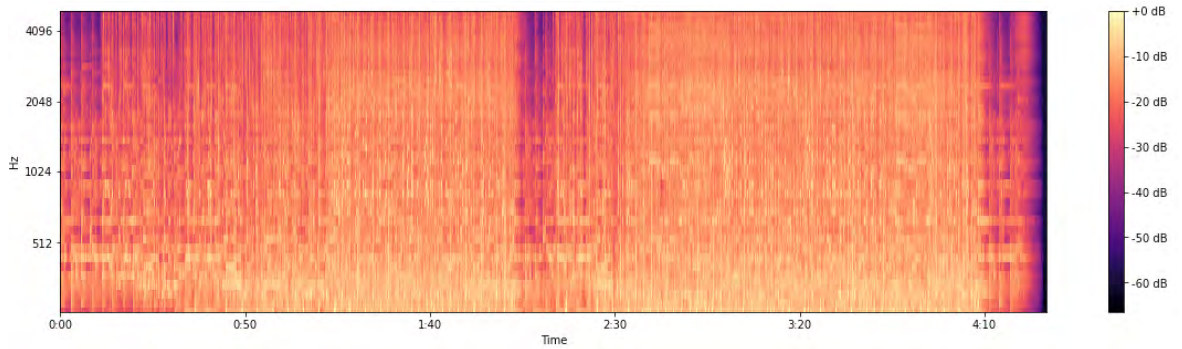


รูปที่ 3.3 รูปแบบของเครื่องหมายกำหนดจังหวะของข้อมูลทั้งหมด

ซึ่งรูปที่ 3.2 แสดงถึงการกระจายตัวของข้อมูลเพลง โดยที่แกน x คือรูปแบบของเครื่องหมายกำหนดจังหวะและแกน y คือจำนวนเพลง

3.3 เมลสเปกโทรแกรม

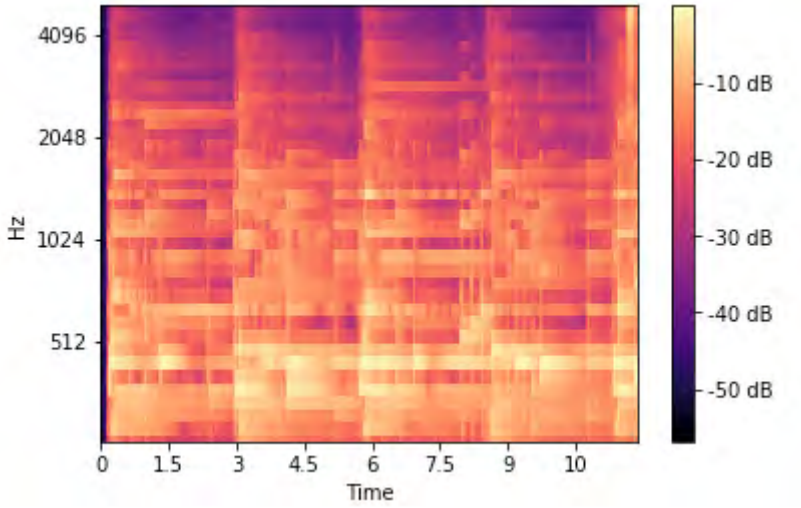
ตัวอย่างเมลสเปกโทรแกรมของเพลง คำยินดี



รูปที่ 3.4 สเปกโทรแกรมของเพลงคำยินดี

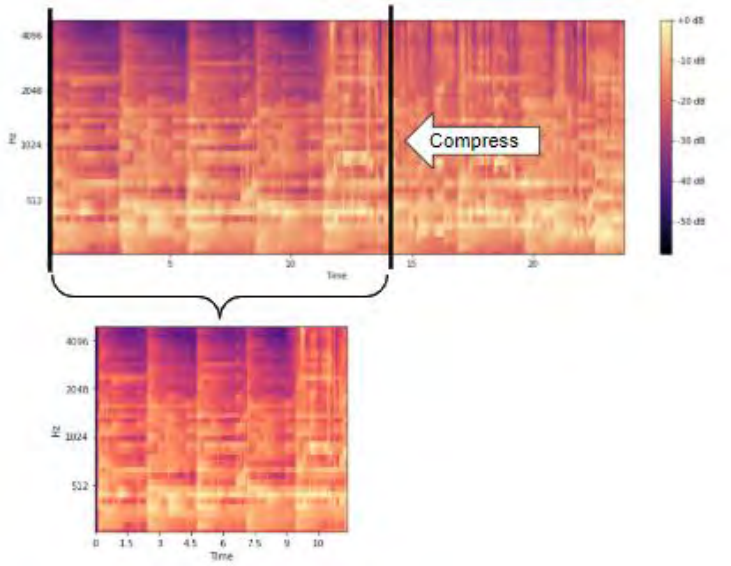
โดยเราสามารถนำเมลสเปกโทรแกรมที่ได้นั้นไปทำการแต่งเติมโดยวิธีการแปลงเลียนแบบ (Apply Affine Transform) ซึ่งเป็นการนำเมลสเปกโทรแกรมมาทำการขยายหรือหดภาพตามแนวแกน x ซึ่งเป็นแกนของเวลา ทำให้เพลงมีความเร็วขึ้นหรือช้าลง อีกทั้งยังปรับป้ายกำกับให้สอดคล้องจะได้เมลสเปกโทรแกรมใหม่ขึ้นมาดังต่อไปนี้

เมลสเปกโทรแกรมที่มีขนาด 256 เฟรมจะเท่ากับเพลงประมาณ 11.9 วินาที



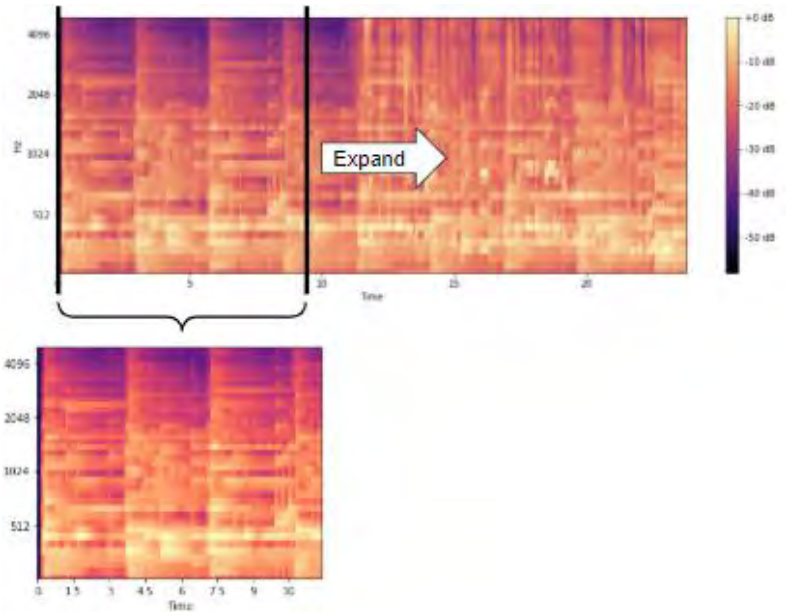
รูปที่ 3.5 เมลสเปกโทรแกรมของเพลงคำยินดี ขนาด 256 เฟรมและไม่มีการแต่งเติม

เมลสเปกโทรแกรมที่มีขนาด 256 เฟรมและมีค่าแต่งเติมเท่ากับ 1.2 จะได้เพลงที่เร็วขึ้นและนำอัตราเร็วจังหวะมาคูณด้วยค่าแต่งเติม เช่น เพลงที่มีอัตราเร็วจังหวะเป็น 100 เมื่อแต่เติมแล้วจะได้เพลงที่มีอัตราเร็ว 120



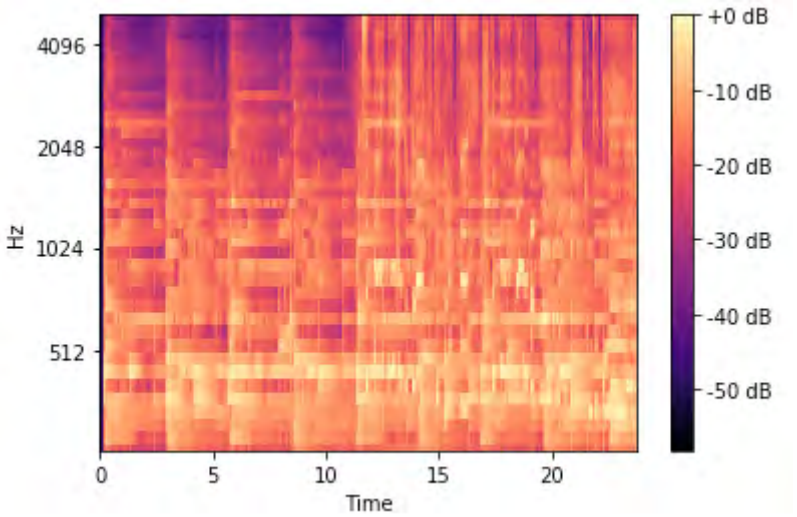
รูปที่ 3.6 สเปกโทรแกรมของเพลงคำยินดี ขนาด 256 เฟรมและมีค่าแต่งเติมเท่ากับ 1.2

เมลสเปกโทรแกรมที่มีขนาด 256 เฟรมและมีค่าแต่งเติมเท่ากับ 0.8 จะได้เพลงที่ช้าลงและนำอัตราเร็วจังหวะมาคูณด้วยค่าแต่งเติม เช่น เพลงที่มีอัตราเร็วจังหวะเป็น 100 เมื่อแต่เติมแล้วจะได้เพลงที่มีอัตราเร็ว 80



รูปที่ 3.7 สเปกโทรแกรมของเพลงคำยินดี ขนาด 256 เฟรมและมีค่าแต่งเติมเท่ากับ 0.8

เมลสเปกโทรแกรมที่มีขนาด 512 เฟรมจะได้เพลงประมาณ 23.8 วินาที



รูปที่ 3.8 สเปกโทรแกรมของเพลง คำยินดี ขนาด 512 เฟรมและไม่มีการแต่งเติม

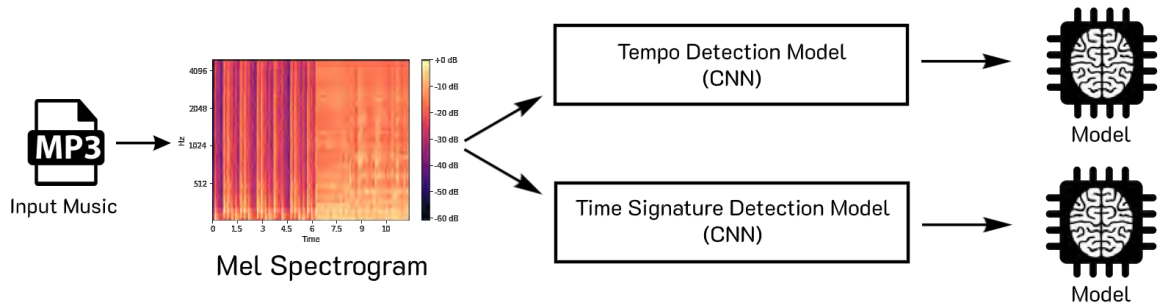
บทที่ 4

ขั้นตอนการวิจัย

จากการศึกษางานวิจัยต่าง ๆ ที่ผ่านมา ทางผู้วิจัยได้นำวิธีการทำงานของงานวิจัยในอดีตมาพัฒนาต่อ โดยการสร้างชั้นภาพ (Layer) ใหม่ขึ้นมาทำการทดลอง โดยคาดหวังว่าผลลัพธ์ที่ได้นั้นจะมีประสิทธิภาพมากขึ้นกว่างานวิจัยที่ผ่านมา

ในบทนี้จะกล่าวถึง รูปแบบของการทำนายอัตราเร็วจังหวะ และเครื่องหมายกำหนดจังหวะ การออกแบบโครงสร้างของโครงข่ายประสาทเทียม วิธีการปรับค่าพารามิเตอร์ และวิธีการทดลอง

4.1รูปแบบของการทำนายอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะ



รูปที่ 4.1 ภาพรวมของระบบ

ภาพรวมของระบบเป็นดังแสดงในรูป 4.1 แบ่งออกเป็น 3 ขั้นตอน

ขั้นตอนแรก เมื่อรับข้อมูลเป็นแฟ้มข้อมูลเอ็มพี 3 มาแล้วจัดเตรียมข้อมูลตามบทที่ 3

ขั้นตอนที่สอง นำสเปกโทรแกรมที่ได้จากการจัดเตรียมข้อมูล เข้าสู่กระบวนการเรียนรู้ของคอมพิวเตอร์ของตัวแบบเพื่อทำนายอัตราเร็วจังหวะของเพลงออกมา

ขั้นตอนที่สาม นำสเปกโทรแกรมที่ได้จากการจัดเตรียมข้อมูล เข้าสู่กระบวนการเรียนรู้ของคอมพิวเตอร์ของตัวแบบเพื่อทำนายเครื่องหมายกำหนดจังหวะของเพลงออกมา

ในขั้นตอนสุดท้ายจึงนำตัวแบบของการทำนายอัตราเร็วจังหวะและตัวแบบของการทำนายเครื่องหมายกำหนดจังหวะ เพื่อนำไปใช้สร้างบริการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะ

4.2 โครงสร้างของโครงข่ายประสาทเทียม

ทางผู้วิจัยได้ลองปรับแต่งโครงสร้างของโครงข่ายประสาทเทียมเพื่อหาโครงสร้างที่เหมาะสมสำหรับการทำนายอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะไว้ดังนี้

4.2.1 โครงสร้างโครงข่ายประสาทเทียมสังวัตนาการสำหรับการทำนายอัตราเร็วจังหวะ

ในส่วนของโครงสร้างโครงข่ายประสาทเทียมแบบสังวัตนาการได้แบ่งออกเป็น 3 ส่วน และมีข้อมูลนำเข้าเป็นเมลสเปกโทรแกรมที่มีขนาด (40, 256, 1) โดย

40 หมายถึง จำนวนช่วงความถี่ในสเกลเมล

256 หมายถึง จำนวนเฟรม ซึ่งเท่ากับเพลง 11.9 วินาที

1 หมายถึง ช่องของสีในสเปกโทรแกรม

4.2.1.1 short filter convolution layers ส่วนนี้มีจุดมุ่งหมายเพื่อจับส่วนเริ่มต้นของเพลงจากสเปกโทรแกรมที่ได้ ประกอบไปด้วย

- Batch Normalization เป็นชั้นที่ช่วยปรับข้อมูลให้มีขนาดที่เหมาะสม ทำให้การเรียนรู้ของตัวแบบเร็วขึ้น
- Convolution layer ขนาด 16x1x5 ที่ใช้ ELU** เป็น Activation function

ทั้งหมด 3 ชุด

4.2.1.2 multi filter modules – ส่วนนี้มีจุดมุ่งหมายเพื่อจับสัญญาณให้ตรงกับตัวกรองชนิดต่าง ๆ โดยแต่ละมอดูลประกอบไปด้วย

- Pooling layer โดยเป็นการ ลดขนาดมิติของข้อมูลแบบเฉลี่ย (Average pooling) เนื่องจากการลดขนาดมิติข้อมูลแบบเฉลี่ยนั้น สามารถเก็บค่าลักษณะสำคัญ เช่น ความเข้มของความถี่ของสเปกโทรแกรม เพื่อนำไปใช้งานต่อได้ ซึ่งหากเป็นการลดขนาดมิติของข้อมูลแบบค่ามากสุด (Max pooling) จะทำให้เกิดการสูญเสียลักษณะสำคัญดังกล่าว ซึ่งจะทำให้ผลลัพธ์ออกมามีค่าความแม่นยำที่น้อยลง
- Batch Normalization เป็นชั้นที่ช่วยปรับข้อมูลให้มีขนาดที่เหมาะสม ทำให้การเรียนรู้ของตัวแบบเร็วขึ้น
- Convolution layer ขนาด 24x1x32
- Convolution layer ขนาด 24x1x64
- Convolution layer ขนาด 24x1x96
- Convolution layer ขนาด 24x1x128

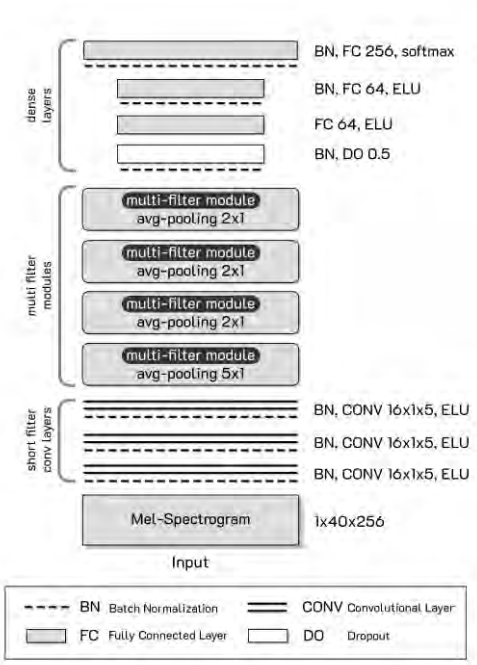
- Convolution layer ขนาด 24x1x192
- Convolution layer ขนาด 24x1x256

โดย Convolution layer จะเชื่อมต่อกันแบบขนาน โดยใช้ Activation function ตัวเดียวกันทั้งหมดคือ ELU**

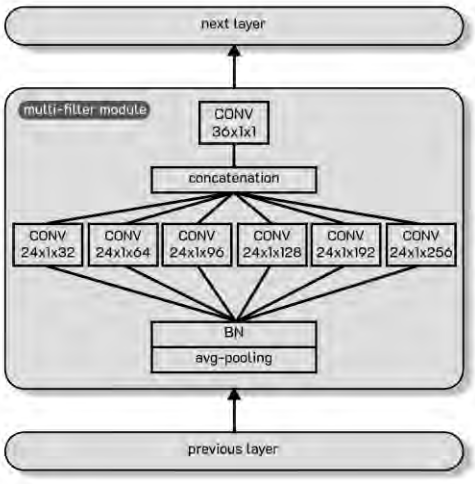
- Concatenation layer เป็นชั้นที่รวม convolution layer ทั้งหมดแล้วส่งข้อมูลออกมา
- Convolution layer ขนาด 36x1x1

4.2.1.3 dense layers - ส่วนนี้มีจุดมุ่งหมายเพื่อแยกประเภทของอัตราเร็วจังหวะ ประกอบไปด้วย

- Batch Normalization เป็นชั้นที่ช่วยปรับข้อมูลให้มีขนาดที่เหมาะสม ทำให้การเรียนรู้ของตัวแบบเร็วขึ้น
- Drop out layer เป็นชั้นที่สุ่มตัดเส้นที่เชื่อมระหว่างชั้นก่อนหน้าไปยังชั้นถัดไปเพื่อป้องกันการ overfitting
- Fully Connected layer ขนาด 64 โดยมี activation function คือ ELU**
- Batch Normalization เป็นชั้นที่ช่วยปรับข้อมูลให้มีขนาดที่เหมาะสม ทำให้การเรียนรู้ของตัวแบบเร็วขึ้น
- Fully Connected layer ขนาด 64 โดยมี activation function คือ ELU**
- Batch Normalization เป็นชั้นที่ช่วยปรับข้อมูลให้มีขนาดที่เหมาะสม ทำให้การเรียนรู้ของตัวแบบเร็วขึ้น
- Fully Connected layer ขนาด 256 โดยมี activation function คือ SoftMax



รูปที่ 4.2 โครงสร้างโครงข่ายประสาทเทียมสังวัตนาการสำหรับการทำนายอัตราเร็วจังหวะที่มีการดัดแปลงจากงานวิจัย [1]



รูปที่ 4.3 multifilter module ในโครงสร้างโครงข่ายประสาทเทียมสังวัตนาการสำหรับการทำนายอัตราเร็วจังหวะที่มีการดัดแปลงจากงานวิจัย [1]

4.2.2 โครงสร้างโครงข่ายประสาทเทียมแบบสังวัตนาการสำหรับการทำนายเครื่องหมายกำหนดจังหวะ

ในส่วนของโครงสร้างโครงข่ายประสาทเทียมแบบสังวัตนาการได้แบ่งออกเป็น 3 ส่วน และมีข้อมูลนำเข้าเป็นเมลสเปกโทรแกรมที่มีขนาด (40, 512, 1) โดย

40 หมายถึง ความถี่ในหน่วยเมลสเกล

512หมายถึง ขนาดของเฟรม ซึ่งเท่ากับเพลง 23.8 วินาที

1 หมายถึง ช่องของสีในสเปกโทรแกรม

4.2.2.1 short filter convolution layers ส่วนนี้มีจุดมุ่งหมายเพื่อจับส่วนเริ่มต้นของเพลงจากสเปกโทรแกรมที่ได้ ประกอบไปด้วย

- Batch Normalization เป็นชั้นที่ช่วยปรับข้อมูลให้มีขนาดที่เหมาะสม ทำให้การเรียนรู้ของตัวแบบเร็วขึ้น
- Convolution layer ขนาด 16x1x5 ที่ใช้ ELU** เป็น Activation function

ทั้งหมด 3 ชุด

4.2.2.2 multi filter modules – ส่วนนี้มีจุดมุ่งหมายเพื่อจับสัญญาณให้ตรงกับตัวกรองชนิดต่าง ๆ โดยแต่ละมอดูลประกอบไปด้วย

- Pooling layer โดยเป็นการ ลดขนาดมิติของข้อมูลแบบเฉลี่ย (Average pooling) เนื่องจากการลดขนาดมิติข้อมูลแบบเฉลี่ยนั้น สามารถเก็บค่าลักษณะสำคัญ เช่น ความเข้มของความถี่ของสเปกโทรแกรม เพื่อนำไปใช้งานต่อได้ ซึ่งหากเป็นการลดขนาดมิติของข้อมูลแบบค่ามากสุด (Max pooling) จะทำให้เกิดการสูญเสียลักษณะสำคัญดังกล่าว ซึ่งจะทำให้ผลลัพธ์ออกมามีค่าความแม่นยำที่น้อยลง
- Batch Normalization เป็นชั้นที่ช่วยปรับข้อมูลให้มีขนาดที่เหมาะสม ทำให้การเรียนรู้ของตัวแบบเร็วขึ้น
- Convolution layer ขนาด 24x1x32
- Convolution layer ขนาด 24x1x64
- Convolution layer ขนาด 24x1x96
- Convolution layer ขนาด 24x1x128

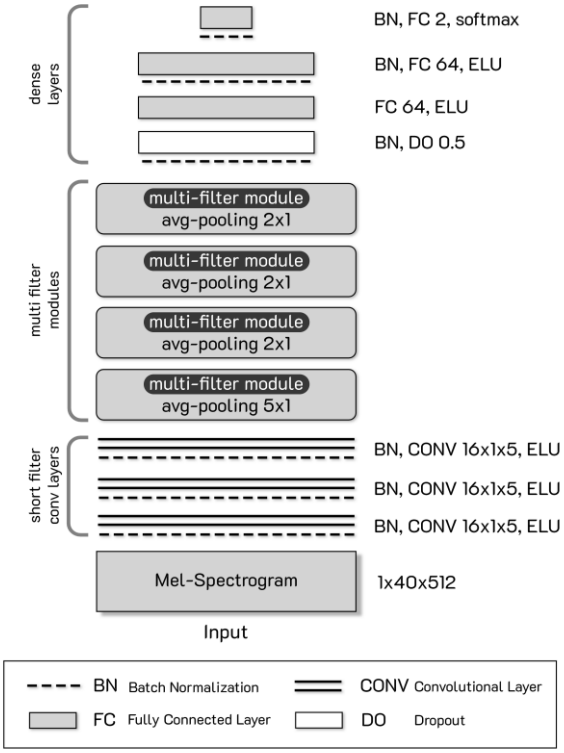
- Convolution layer ขนาด 24x1x192
- Convolution layer ขนาด 24x1x256
- Convolution layer ขนาด 24x1x512

โดย Convolution layer จะเชื่อมต่อกันแบบขนาน โดยใช้ Activation function ตัวเดียวกันทั้งหมดคือ ELU**

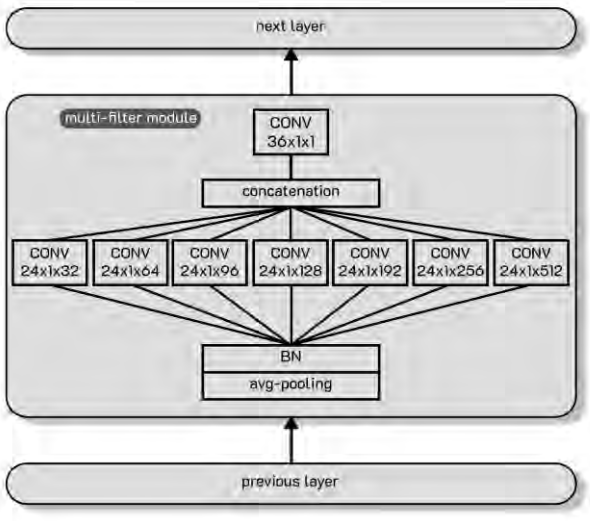
- Concatenation layer เป็นชั้นที่รวม convolution layer ทั้งหมดแล้วส่งข้อมูลออกมา
- Convolution layer ขนาด 36x1x1

4.2.2.3 dense layers - ส่วนนี้มีจุดมุ่งหมายเพื่อแยกประเภทของเครื่องหมายกำหนดจังหวะ ประกอบไปด้วย

- Batch Normalization เป็นชั้นที่ช่วยปรับข้อมูลให้มีขนาดที่เหมาะสม ทำให้การเรียนรู้ของตัวแบบเร็วขึ้น
- Drop out layer – เป็นชั้นที่สุ่มตัดเส้นที่เชื่อมระหว่างชั้นก่อนหน้าไปยังชั้นถัดไปเพื่อป้องกันการ overfitting
- Fully Connected layer ขนาด 64 โดยมี activation function คือ ELU**
- Batch Normalization เป็นชั้นที่ช่วยปรับข้อมูลให้มีขนาดที่เหมาะสม ทำให้การเรียนรู้ของตัวแบบเร็วขึ้น
- Fully Connected layer ขนาด 64 โดยมี activation function คือ ELU**
- Batch Normalization เป็นชั้นที่ช่วยปรับข้อมูลให้มีขนาดที่เหมาะสม ทำให้การเรียนรู้ของตัวแบบเร็วขึ้น
- Fully Connected layer ขนาด 2 โดยมี activation function คือ softmax



รูปที่ 4.4 โครงสร้างโครงข่ายประสาทเทียมสังวัตนาการสำหรับการทำนายเครื่องหมายกำหนดจังหวะที่มีการดัดแปลงจากงานวิจัย [1]



รูปที่ 4.5 multifilter module ในโครงสร้างโครงข่ายประสาทเทียมสังวัตนาการสำหรับการทำนายเครื่องหมายกำหนดจังหวะที่มีการดัดแปลงจากงานวิจัย [1]

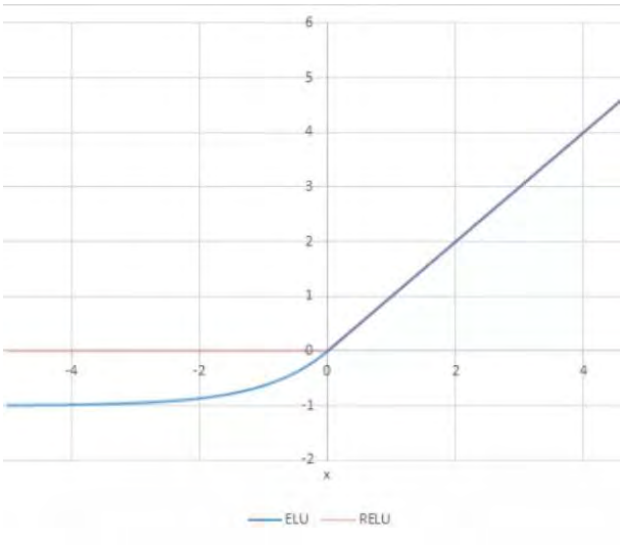
**ELU หรือ Exponential Linear Unit เป็น activation function ชนิดหนึ่งซึ่งต่างจาก activation function ชนิดอื่นเพราะมีตัวแปรอัลฟา (α) โดยจะมีสมการดังนี้

$$R(z) = \begin{cases} z & z > 0 \\ \alpha(e^z - 1) & z \leq 0 \end{cases}$$

โดย z คือค่านำเข้า

α คือ ค่าตัวแปรเสริมขั้นสูง (Hyperparameter)

e คือ ค่าคงตัวทางคณิตศาสตร์ ที่เป็นฐานของลอการิทึมธรรมชาติ (ค่าประมาณ 2.71828)



รูปที่ 4.6 กราฟของ ELU

ที่มา: https://ml-cheatsheet.readthedocs.io/en/latest/activation_functions.html

โดยหากเราทำการใส่ค่า z มากกว่า 0 ผลลัพธ์ที่ได้ก็จะเท่ากับค่า z แต่หากเราใส่ค่า z น้อยกว่าหรือเท่ากับ 0 ผลลัพธ์ที่ได้จะขึ้นอยู่กับค่า z และค่าอัลฟา ซึ่งเราสามารถปรับค่าอัลฟานี้ได้

4.3 การทดลอง

4.3.1 การทดลองตัวแบบสำหรับอัตราเร็วจังหวะ

ตัวแบบนี้จะเป็นการทำนายแบบจำแนกประเภทเป็น 256 ประเภท ตั้งแต่ 30 – 286 ซึ่งเป็นค่าอัตราเร็วจังหวะทั้งหมด 256 ค่าที่แตกต่างกันโดยตัวแบบที่ใช้จะใช้เป็นตัวแบบที่ได้รับการฝึกสอนมาแล้วจากงานวิจัยที่ได้ทำการศึกษาซึ่งทางผู้วิจัยได้มีการทดลองต่าง ๆ ดังนี้

4.3.1.1 การใส่ข้อมูลนำเข้าเป็นชุดข้อมูลสำหรับให้คอมพิวเตอร์เรียนรู้เตรียมไว้ โดยจะทำการปรับน้ำหนักของชั้นทั้งหมดให้เข้ากับชุดข้อมูลที่ใส่เข้าไป

4.3.1.2 การทดลองลบค่าน้ำหนักของ 5 ชั้นสุดท้าย และมีการล็อคน้ำหนักของชั้นก่อนหน้าทั้งหมด จากนั้นจึงใส่ข้อมูลที่เป็นชุดข้อมูลที่สำหรับให้คอมพิวเตอร์เรียนรู้

4.3.1.3 การทดลองสร้างตัวแบบขึ้นใหม่ โดยมีลักษณะโครงสร้างคล้ายเดิม แต่ไม่มีค่าน้ำหนักอยู่ในแต่ละชั้น

ในแต่ละการทดลองจะมีการปรับเปลี่ยนค่าอัตราการเรียนรู้ (learning rate) โดยมีค่าดังนี้ 0.001 0.005 0.01 0.05 รวมถึงมีการปรับเปลี่ยน optimizer ให้เป็น Adam หรือ Stochastic Gradient Descent และใช้ early stopping เพื่อทำการหยุดการเรียนรู้เมื่อค่า validation lost สูงขึ้นเกินกว่าที่กำหนดไว้

4.3.2 การทดลองตัวแบบสำหรับเครื่องหมายกำหนดจังหวะ

ตัวแบบนี้จะเป็นการทำนายแบบจำแนกประเภทเป็น 2 ประเภท คืออัตราจังหวะ 4/4 และ 6/8 โดยปรับจากโครงสร้างตัวแบบของอัตราเร็วจังหวะ โดยหาอัตราเร็วจังหวะแต่รับค่าเป็น (40,512,1) แล้วทำการเพิ่มชั้น convolutional filter และมีการปรับชั้นสุดท้ายให้เหลือตัวจำแนกประเภทอยู่ 2 ตัวเลือกคือ 4/4 หรือ 6/8

4.3.2.1 การใส่ข้อมูลนำเข้าเป็นชุดข้อมูลที่สำหรับให้คอมพิวเตอร์เรียนรู้เตรียมไว้ โดยจะทำการปรับน้ำหนักของชั้นทั้งหมดให้เข้ากับชุดข้อมูลที่เตรียมไว้

4.3.2.2 การทดลองสร้างตัวแบบขึ้นใหม่ โดยมีลักษณะโครงสร้างคล้ายเดิม แต่ไม่มีค่าน้ำหนักอยู่ในแต่ละชั้น

ในแต่ละการทดลองจะใช้ค่าอัตราการเรียนรู้เท่ากับ 0.005 รวมถึงมีการใช้ Adam เป็น optimizer และใช้ early stopping เพื่อทำการหยุดการเรียนรู้เมื่อค่า validation lost สูงขึ้นเกินกว่าที่กำหนดไว้

4.4 การวัดผลการวิจัย

4.4.1 การวัดผลการวิจัยของอัตราเร็วจังหวะ

วัดผลโดยค่าความแม่นยำ (accuracy) 3 ค่าโดยเฉลี่ยจากการทดสอบตัวแบบด้วยชุดข้อมูลทดสอบทั้งหมด 10 ชุด

Accuracy0 - ค่าความแม่นยำที่เกิดจากการทำนายอัตราเร็วจังหวะแบบถูกร้อยละ 100

Accuracy1 - ค่าความแม่นยำที่เกิดจากการทำนายอัตราเร็วจังหวะแบบยอมรับค่าความคลาดเคลื่อนร้อยละ 4

Accuracy2 - ค่าความแม่นยำที่เกิดจากการทำนายอัตราเร็วจังหวะแบบยอมรับค่าความคลาดเคลื่อนร้อยละ 4 และยอมรับความคลาดเคลื่อนในอัตราส่วน 1/2 1/3 2 3 เท่าของอัตราเร็วจังหวะ

4.4.2 การวัดผลการวิจัยของเครื่องหมายกำหนดจังหวะ

วัดผลโดยค่า f1-score จากการทำนายค่าของเครื่องหมายกำหนดจังหวะแบบ 4/4 และ 6/8 ซึ่งเป็นค่าเฉลี่ยของค่าความแม่นยำ (precision) และค่าความครบถ้วน (recall) และยังช่วยวัดผลข้อมูลโดยที่ข้อมูลมีความไม่สมดุล จากการทดสอบตัวแบบด้วยชุดข้อมูลทดสอบทั้งหมด 10 ชุด

4.5 การพัฒนาบริการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะ

การพัฒนาบริการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะนั้น จะเป็นการนำตัวแบบที่มีประสิทธิภาพมากที่สุดในการทำนายอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะไปพัฒนาเป็นบริการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะตามขั้นตอนดังนี้

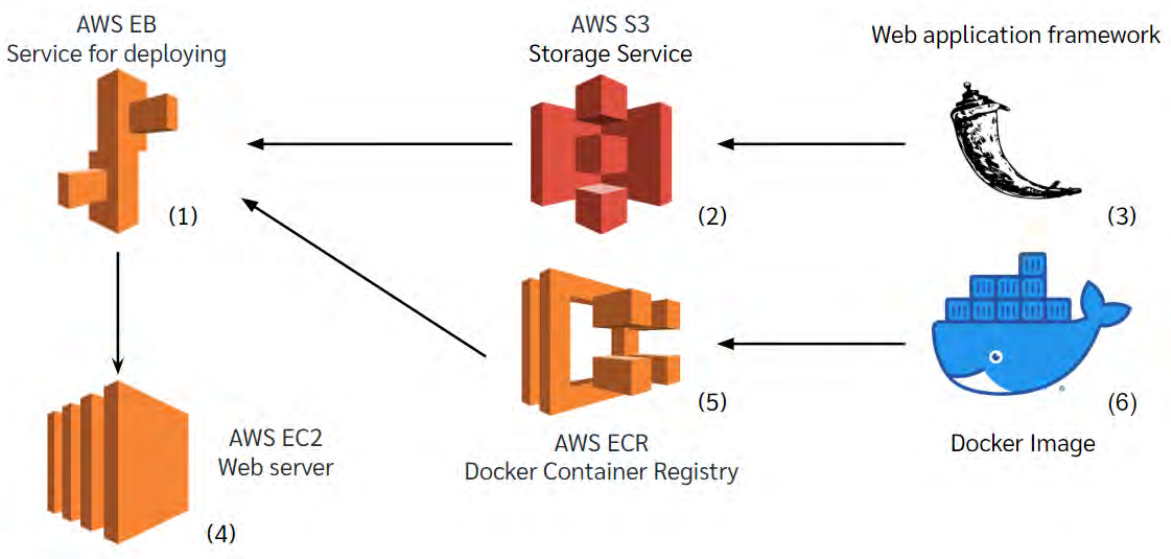
4.5.1 สร้างเอพีไอด้วยเฟรมเวิร์ค Flask

4.5.2 สร้าง Docker image เพื่อลงโปรแกรมสนับสนุนต่าง ๆ ที่จำเป็นในการใช้ร่วมกับบริการ

4.5.3 นำอิมเมจของ Docker ที่สร้างฝากไว้ที่ Amazon ECR และนำส่วนของโค้ดฝากไว้ที่ Amazon S3

4.5.4 นำส่วนโค้ด และอิมเมจเข้าสู่บริการออนไลน์ด้วย Amazon Elastic Beanstalk เป็นผู้จัดการให้

ซึ่งสามารถแสดงการทำงานได้ดังรูปที่ 4.7 และคู่มือการใช้งานจะแสดงดังภาคผนวก ก



รูปที่ 4.7 ภาพรวมระบบบริการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะ

ที่มา: (1) https://aws.amazon.com/elasticbeanstalk/

(2) https://www.zabbix.com/integrations/aws_s3

(3) https://seeklogo.com/vector-logo/273085/flask

(4) https://www.zabbix.com/integrations/aws_ec2

(5) https://medium.com/rapids-ai/getting-started-with-rapids-on-aws-ecs-using-dask-cloud-provider-b1adfdbc9c6e

(6) https://www.blognone.com/node/113131

บทที่ 5

ผลการวิจัย

ในบทนี้จะกล่าวถึง ผลการดำเนินการวิจัยของการทำนายอัตราเร็วจังหวะและผลการดำเนินการวิจัยของการทำนายเครื่องหมายกำหนดจังหวะด้วยโมเดลหลากหลายแบบดังนี้

5.1 ผลการดำเนินการวิจัยของการทำนายอัตราเร็วจังหวะด้วยโมเดลทั้ง 7 แบบ

ตารางที่ 5.1 สรุปตัวแบบต่าง ๆ จากค่าความถูกต้องเฉลี่ย

	Accuracy0 (%)	Accuracy1 (%)	Accuracy2 (%)
Model1	61.25	70.00	96.00
Model2	**73.25**	**78.00**	**98.75**
Model3	57.78	70.25	95.25
Model4	65.00	72.50	99.00
Model5	71.25	76.25	99.50
Model6	69.75	77.50	99.75
Model7	64.50	73.50	99.00

Model1 คือ ตัวแบบที่นำมาจากงานวิจัย [1]

Model2 คือ ตัวแบบที่นำมาจากงานวิจัย [1] และปรับน้ำหนักตามข้อมูลนำเข้า โดยมีค่าอัตราการเรียนรู้เป็น 0.001 ด้วย optimizer Adam

Model3 คือ ตัวแบบที่นำมาจากงานวิจัย [1] และปรับน้ำหนักของชั้น 5 ชั้นสุดท้าย รวมถึงล็อคน้ำหนักของชั้นที่เหลือ

Model4 คือ ตัวแบบที่นำมาจากงานวิจัย [1] และปรับน้ำหนักตามข้อมูลนำเข้า โดยมีค่าอัตราการเรียนรู้เป็น 0.005

Model5 คือ ตัวแบบที่นำมาจากงานวิจัย [1] และปรับน้ำหนักตามข้อมูลนำเข้า โดยมีค่าอัตราการเรียนรู้เป็น 0.05 ด้วย optimizer SGD

Model6 คือ ตัวแบบที่นำมาจากงานวิจัย [1] และปรับน้ำหนักตามข้อมูลนำเข้า โดยมีค่าอัตราการเรียนรู้เป็น 0.1 ด้วย optimizer SGD

Model7 คือ ตัวแบบที่นำมาจากงานวิจัย [1] และปรับน้ำหนักทุกชั้นใหม่ โดยมีค่าอัตราเรียนรู้เป็น 0.001 ด้วย optimizer Adam

5.2 ผลการดำเนินการวิจัยของการทำนายเครื่องหมายกำหนดจังหวะด้วยโมเดลทั้ง 2 แบบ

ตารางที่ 5.2 สรุปค่าความแม่นยำ (Precision) ความครบถ้วน (Recall) และ F1-score ของตัวแบบที่ 1

Model 1			
	Precision (%)	**Recall (%)**	**f1-score (%)**
4-4	**95.40**	**75.70**	**79.80**
6-8	**11.40**	**43.40**	**15.00**

ตารางที่ 5.3 สรุปค่าความแม่นยำ (Precision) ความครบถ้วน (Recall) และ F1-score ของตัวแบบที่ 2

Model2			
	Precision (%)	**Recall (%)**	**f1-score (%)**
4-4	**83.60**	**61.70**	**64.70**
6-8	**31.11**	**26.56**	**11.80**

ตารางที่ 5.4 สรุปค่าความแม่นยำของแต่ละตัวแบบ

	Accuracy(%)
Model1	73.50
Model2	57.90

Model1 คือ ตัวแบบที่นำมาจากงานวิจัย [1] และปรับน้ำหนักทุกชั้นใหม่ โดยมีค่าอัตราการเรียนรู้เป็น 0.05 ด้วย optimizer Adam

Model2 คือ ตัวแบบที่นำมาจากงานวิจัย [1] และลบชั้น conv 16x1x32 ใน multifilter module ออก โดยมีค่าอัตราการเรียนรู้เป็น 0.05 ด้วย optimizer Adam

5.3 สรุปผลของการดำเนินการวิจัย

จากผลการดำเนินการวิจัยของการทำนายอัตราเร็วจังหวะ จะสังเกตได้ว่า Model2 ซึ่งเป็นตัวแบบที่นำมาจากงานวิจัย [1] และปรับน้ำหนักตามข้อมูลนำเข้า โดยมีค่าอัตราการเรียนรู้เป็น 0.001 ด้วย optimizer Adam มีค่าความถูกต้องเฉลี่ยโดยรวมสูงที่สุดเทียบกับตัวแบบทั้งหมด เมื่อนำไปเปรียบเทียบกับ Model1 ที่ได้รับการเรียนรู้มาจากชุดข้อมูลที่เกี่ยวกับอัตราเร็วจังหวะ จะเห็นได้ว่า มีค่าความถูกต้องเฉลี่ยสูงขึ้น เนื่องจากได้รับการเรียนรู้มาจากข้อมูลที่ทางผู้วิจัยได้เพิ่มเข้าไป ทำให้มีความสามารถในการทำนายที่สูงขึ้น โดยจะแตกต่างจาก Model7 ซึ่งเป็นตัวแบบที่ทำการปรับค่าน้ำหนักใหม่ทั้งหมด ทำให้เกิดการเรียนรู้เฉพาะข้อมูลที่ทางผู้วิจัยได้เพิ่มเข้าไป ส่งผลให้มีข้อมูลสำหรับการเรียนรู้ที่น้อย ความสามารถในการทำนายจึงน้อยลงตามมาด้วย และเมื่อนำไปเปรียบเทียบกับ Model4 Model5 และ Model6 ที่มีการปรับแต่งค่าอัตราการเรียนรู้ หรือการเปลี่ยนตัว optimizer จะเห็นได้ว่า ค่าความถูกต้องเฉลี่ยน้อยลง จึงสามารถสรุปได้ว่าการปรับค่าต่าง ๆ ไม่เหมาะสมสำหรับการทำนายอัตราเร็วจังหวะ

จากผลการดำเนินการวิจัยของการทำนายเครื่องหมายกำหนดจังหวะ จะสังเกตได้ว่า Model1 มีค่า f1-score ที่สูงกว่า Model2 ที่มีการนำชั้นใน multifilter module ออกจึงทำให้ได้ข้อสรุปว่า การปรับแต่งตัวแบบโดยการนำชั้นใน multifilter module ออกส่งผลให้ความสามารถในการทำนายลดน้อยลงตามไปด้วย

สำหรับตัวแบบในการทำนายเครื่องหมายกำหนดจังหวะทั้ง 2 ตัวแบบ ทางผู้วิจัยได้พบว่า ทุกตัวแบบนั้นเกิดปัญหาเนื่องจากตัวแบบไม่สามารถแยกเครื่องหมายกำหนดจังหวะ 6/8 ออกจากเครื่องหมายกำหนดจังหวะ 4/4 ได้ซึ่งอาจจะเกิดจากชุดข้อมูลที่มีน้อยมากเกินไป จนทำให้ตัวแบบไม่สามารถเรียนรู้เครื่องหมายกำหนดจังหวะแบบ 6/8 ได้เพียงพอ เนื่องจากเพลงไทยประเภทพอป ร็อกแอนด์โรล และ อาร์แอนด์บี ที่มีเครื่องหมายกำหนดจังหวะ 6/8 นั้นมีจำนวนน้อยมากตามธรรมชาติ

บทที่ 6

ข้อสรุปและข้อเสนอแนะ

6.1 สรุปผลการดำเนินงาน

สำหรับการทำนายอัตราเร็วจังหวะในงานนี้ ทางผู้วิจัยได้ทดลองปรับตัวแบบด้วยวิธีการปรับน้ำหนักของชั้น การปรับค่าอัตราการเรียนรู้ การเปลี่ยนตัว optimizer ผลการทดลองในด้านความถูกต้อง มีผลออกมาน่าพึงพอใจ

สำหรับการทำนายเครื่องหมายกำหนดจังหวะ ทางผู้วิจัยได้ทดลองปรับตัวแบบด้วยวิธีการปรับน้ำหนักของชั้นและมีการลบชั้นใน multifilter module ออก ผลการทดลองในด้านความถูกต้อง มีผลออกมาไม่เป็นที่น่าพึงพอใจ เนื่องจากตัวแบบสามารถทำนายเครื่องหมายกำหนดจังหวะแบบ 6/8 ออกมามีความถูกต้องน้อย

6.2 ปัญหาของงานวิจัยและวิธีการแก้ไข

ปัญหาที่ 1: ป้ายกำกับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะของ Spotify ที่ได้มาจากการใช้อัลกอริทึมทำให้ได้ค่าไม่ค่อยตรง

วิธีแก้ไขปัญหา: ทางผู้วิจัยได้ทำการนับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะแล้วสร้างเป็นป้ายกำกับเอง

ปัญหาที่ 2: ข้อมูลเพลงที่นำมาใช้มีจำนวนที่น้อย

วิธีแก้ไขปัญหา: นำข้อมูลเพลงทั้งหมดไปทำการแต่งเติมโดยค่าแต่งเติมที่มีค่าตั้งแต่ 0.8 ถึง 1.2 เพื่อให้ได้ข้อมูลออกมาทั้งหมดประมาณ 1000 ข้อมูล

ปัญหาที่ 3: โครงสร้างของโครงข่ายประสาทเทียมแบบสังวัตนาการที่มีข้อมูลนำเข้าเป็นเมลสเปกโทรแกรมที่มีขนาด (40, 512, 1) ซึ่งเป็นโครงข่ายประสาทเทียมที่มีขนาดใหญ่นั้นถือเป็นข้อจำกัดอย่างหนึ่งที่ไม่สามารถทำให้การสอนตัวแบบมีประสิทธิภาพมาก เพราะต้องใช้ทรัพยากรของเครื่องเป็นจำนวนมากในการสอนตัวแบบนั้น ๆ และใช้เวลาในการฝึกสอนนาน อีกทั้งการใช้ทรัพยากรของเครื่องเป็นจำนวนมากสำหรับ AWS SageMaker แล้ว ทำให้มีค่าใช้จ่ายที่สูง

วิธีแก้ไขปัญหา: เลือกใช้ AWS Sagemaker ที่มีประสิทธิภาพ แต่แลกมาด้วยค่าใช้จ่ายที่สูงตามมา

รายการอ้างอิง

[1] Hendrik Schreiber, Meinard Müller. A Single-Step Approach To Musical Tempo Estimation Using A Convolutional Neural Network. Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR) (2561) : 98-105

[2] Jan Schlüter, Sebastian Böck. Musical Onset Detection with Convolutional Neural Networks. 6th International Workshop on Machine Learning and Music (MML) (2556)

[3] Brian McFee, Colin Raffel, Dawen Liang, Daniel P.W. Ellis, Matt McVicar, Eric Battenbergk, Oriol Nieto. LibROSA: Audio and Music Signal Analysis in Python. PROC. OF THE 14th PYTHON IN SCIENCE CON. SCIPY (2558)

[4] James Robert. Pydub.[ระบบออนไลน์]. แหล่งที่มา : https://github.com/jiaaro/pydub [1 กันยายน 2562]

[5] Brian McFee, Stefan Balke. LibROSA.[ระบบออนไลน์] แหล่งที่มา : https://LibROSA.github.io/LibROSA/ [1 กันยายน 2562]

[6] สำเร็จ คำโมง. ทฤษฎีดนตรีสากลฉบับสรรพสูตร. กรุงเทพมหานคร : สำนักพิมพ์ฐานบัณฑิต. 2553

[7] สมศักดิ์ สร้อยระย้า. จังหวะ (The Rhythm). กรุงเทพมหานคร : สำนักพิมพ์โอเดียนสโตร, 2545.

[8] ณัชชา พันธุ์เจริญ. ทฤษฎีดนตรี. พิมพ์ครั้งที่ 13. กรุงเทพมหานคร : สำนักพิมพ์เกศกะรัต, 2558.

[9] ศานติ เดชคำรณ. เอกสารประกอบการสอน ทฤษฎีดนตรีสากล 1. เอกสารประกอบการสอน ทฤษฎีดนตรีสากล 1 คณะศิลปกรรมศาสตร์ : มหาวิทยาลัยบูรพา, 2548.

ภาคผนวก

ภาคผนวก ก

คู่มือการใช้งาน

การใช้งานบริการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะ

ตัวอย่างโปรแกรมภาษาไพธอนที่เขียน request ประเภท POST โดยจะมีการส่งแฟ้มข้อมูลของเพลงไปที่ยูอาร์แอล (Universal Resource Locator: URL) เพื่อเรียกใช้บริการดังรูปที่ ก.1 และจะได้ผลลัพธ์ออกมาเป็น JSON ที่มีค่าแสดงดังรูปที่ ก.2

```python
# Install the Python Requests library:
# `pip install requests`

import requests


def send_request():
    # Tempo Meter
    # POST http://tempometer-service.ap-southeast-1.elasticbeanstalk.com/tempo

    try:
        response = requests.post(
            url="http://tempometer-service.ap-southeast-1.elasticbeanstalk.com/tempo",
            headers={
                "Content-Type": "multipart/form-data; charset=utf-8; boundary=__X_PAW_BOUNDARY__",
            },
            files={
                "file": "",
            },
        )
        print('Response HTTP Status Code: {status_code}'.format(
            status_code=response.status_code))
        print('Response HTTP Response Body: {content}'.format(
            content=response.content))
    except requests.exceptions.RequestException:
        print('HTTP Request failed')
```

รูปที่ ก.1 ตัวอย่างการเขียนโปรแกรมเพื่อใช้งานบริการเพื่อการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะของเพลงจากแฟ้มข้อมูลเอ็มพี 3


```
{
    "song":"1 2 3 4 5 I love you_The Bottom Blues",
    "tempo":118,
    "time signature":"4/4"
}
```


รูปที่ ก.2 ผลลัพธ์ที่ได้จากการเรียกใช้งานบริการเพื่อการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะของเพลงจากแฟ้มข้อมูลเอ็มพี 3

ภาคผนวก ข

แบบเสนอหัวข้อโครงงาน รายวิชา 2301399 Project Proposal
ปีการศึกษา 2562

ชื่อโครงงาน (ภาษาไทย) บริการเพื่อการตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะของเพลงจากแฟ้มข้อมูลเอ็มพี 3

ชื่อโครงงาน (ภาษาอังกฤษ) Tempo and Time Signature detection service from Mp3

อาจารย์ที่ปรึกษา ผศ.ดร. ฑิตยา หวานวารี

ผู้ดำเนินการ 1. นาย พชระ เจริญกิจ เลขประจำตัวนิสิต 5933640323

2. นาย สุรเกียรติ รอบคอบ เลขประจำตัวนิสิต 5933664423

สาขาวิชาวิทยาการคอมพิวเตอร์

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

หลักการและเหตุผล

ดนตรีนั้นถือกำเนิดขึ้นมาอยู่กับมนุษย์มานานหลากหลายยุคหลายสมัย ซึ่งเป็นสิ่งที่คอยช่วยให้มนุษย์นั้นเกิดความผ่อนคลายหรือช่วยให้เกิดความรื่นเริงบันเทิงใจอย่างมากมาย และมีบุคคลบางกลุ่มที่ต้องการแกะโน้ตดนตรีและทำนองในบทเพลงต่าง ๆ เพื่อให้สามารถเล่นเลียนแบบต้นฉบับได้อย่างถูกต้อง

ส่วนที่สำคัญของดนตรีนั้นคือทำนองและจังหวะของดนตรี ซึ่งสองส่วนนี้เป็นส่วนหลักที่ทำให้สามารถแบ่งบทเพลงออกได้หลายประเภท หลายอารมณ์เพลง หรือหากต้องการแยกประเภทของบทเพลงต่าง ๆ ว่าสามารถจัดให้อยู่ในหมวดหมู่อะไรได้บ้างนั้น จะต้องทราบทำนองและจังหวะของดนตรีในบทเพลงนั้น ๆ ก่อน หลังจากนั้นจึงให้โปรแกรมคอมพิวเตอร์คำนวณออกมาว่าบทเพลงดังกล่าวนั้นอยู่ในหมวดหมู่อะไร

การแกะทำนองและจังหวะของบทเพลงต่าง ๆ นั้นต้องอาศัยความรู้ทางด้านดนตรีหลายด้าน และต้องตรวจจับจังหวะเคาะของตัวโน้ตในแต่ละห้อง และนำข้อมูลเหล่านั้นมารวมกัน หรือตรวจจับตั้งแต่จังหวะเคาะของโน้ตเพลงตัวแรกถึงโน้ตเพลงตัวสุดท้าย หลังจากนั้นจึงพิจารณาจากจังหวะเคาะที่ตรวจจับได้ บริการตรวจจับอัตราเร็วจังหวะนี้จะรับข้อมูลบทเพลงมาในรูปแบบของแฟ้มข้อมูลเอ็มพี 3 (mp3) แล้วแปลงแฟ้มข้อมูลให้อยู่ในรูปแบบของ สเปกโตรแกรม (Spectrogram) ซึ่งเมื่อวิเคราะห์สเปกโตรแกรมแล้ว จะได้ข้อมูลในรูปแบบ JSON โดยข้อมูลที่ได้ออกมานั้นจะเป็นข้อมูลของ อัตราเร็วจังหวะ (Tempo) และเครื่องหมายกำหนดจังหวะ (Time Signature) ของเพลงนั้นๆ

วัตถุประสงค์

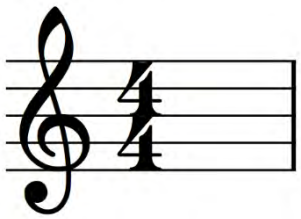
พัฒนาบริการสำหรับตรวจจับอัตราเร็วจังหวะและเครื่องหมายกำหนดจังหวะจากแฟ้มข้อมูลเอ็มพี 3

ความรู้ที่เกี่ยวข้อง

ศานติ เดชคำรณ นิยามศัพท์จังหวะเคาะ อัตราเร็วจังหวะ และอัตราจังหวะ ไว้ใน [9] ดังนี้

- **จังหวะเคาะ (Beat)** คือ หน่วยนับของจังหวะที่ดำเนินไปอย่างสม่ำเสมอ เพื่อที่จะเป็นโครงสร้างให้กับจังหวะประเภทอื่น ๆ เช่น อัตราเร็วจังหวะ อัตราจังหวะ ทำนองเชิงจังหวะ จังหวะประพันธ์ เป็นต้น
- **อัตราเร็วจังหวะ (Tempo)** คือ ค่าอัตราความสัมพันธ์ระหว่างจังหวะเคาะกับเวลา โดยมีหน่วยเป็นจังหวะเคาะต่อนาที หรือ beats per minute (bpm) หากค่าจังหวะเคาะต่อนาทีมีค่าน้อยก็ส่งผลให้เพลงนั้นเป็นเพลงช้า ในทางตรงกันข้ามถ้าหากค่าหากจังหวะเคาะต่อนาทีมีค่ามากก็ส่งผลให้เพลงเป็นเป็นเพลงเร็ว
- **อัตราจังหวะ (Time)** คือ แบบแผนของจังหวะเคาะ โดยอัตราจังหวะแบ่งเป็น 3 ประเภท ได้แก่ อัตราจังหวะธรรมดา อัตราจังหวะผสม อัตราจังหวะซับซ้อน โดยอัตราจังหวะธรรมดาและอัตราจังหวะผสมสามารถพบได้ในเพลงปกติทั่วไป อีกทั้งในเพลงที่นำมาใช้ในการตรวจจับอัตราเร็วของเสียงเพลงนั้นจะเป็นเพลงที่มีสัดส่วนอัตราจังหวะธรรมดา ตามมาด้วยอัตราจังหวะผสม และสุดท้ายด้วยอัตราจังหวะซ้อน

สำเร็จ คำโมง นิยามเครื่องหมายกำหนดจังหวะไว้ใน [6] ดังนี้

- **เครื่องหมายกำหนดจังหวะ (Time Signature)** คือ เครื่องหมายที่เป็นตัวกำหนดจังหวะเคาะในแต่ละห้อง โดยเครื่องหมายชนิดนี้บังคับให้ห้องเพลงแต่ละห้องมีจำนวนนับจังหวะเคาะเท่ากันอย่างตายตัว ลักษณะของเครื่องหมายจะเป็นตัวเลข 2 ตัวเขียนซ้อนกัน คือให้ตัวหนึ่งวางอยู่ข้างบน อีกตัวหนึ่งโดยที่ไม่มีขีดแบ่งดังรูปที่ 1 หรือถ้าเป็นการเขียนบรรยายก็จะใช้เครื่องหมาย / คั่น เช่น 4/4, 6/8

รูปที่ 1

การอ่านเครื่องหมายกำหนดจังหวะแบบอัตราจังหวะธรรมดานั้น เลขตัวบนบอกจำนวนจังหวะเคาะได้ทันที และเลขตัวล่างก็บอกถึงรูปร่างของตัวโน้ตที่ใช้นับเป็น 1 จังหวะเคาะได้ทันทีโดยเลขตัวเลขสื่อความหมายดังนี้

เลข 1 ให้นับโน้ตตัวกลม (𝅝) จังหวะเคาะ

เลข 2 ให้นับโน้ตขาว (𝅗𝅥) เป็น 1 จังหวะเคาะ

เลข 4 ให้นับโน้ตตัวดำ (♩) เป็น 1 จังหวะเคาะ

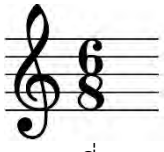
เลข 8 ให้นับโน้ตตัวเขบ็ตชั้นเดียว (♪) เป็น 1 จังหวะเคาะ

เลข 16 ให้นับโน้ตตัวเขบ็ตสองชั้น (♬) เป็น 1 จังหวะเคาะ

และเครื่องหมายกำหนดจังหวะแบบอัตราจังหวะผสม ต้องตีความหมายของตัวเลขทั้งบนและล่างก่อนถึงจึงจะนับจำนวนจังหวะเคาะได้ โดย

เอา 3 หาร เลขตัวบน ผลลัพธ์จะเป็นจำนวนจังหวะเคาะต่อ 1 ห้อง

เอา 2 หาร เลขตัวล่าง ผลลัพธ์จะเป็นตัวโน้ตรูปร่างใดประจุด เป็น 1 จังหวะเคาะ

ตัวอย่างเช่น เครื่องหมายกำหนดจังหวะ 6/8 ดังรูปที่ 2

รูปที่ 2

นำ 6 ÷ 3 = 4 จะได้ว่า 1 ห้องมี 2 จังหวะ

นำ 8 ÷ 2 = 4 จะได้ว่านับเอาโน้ตตัวดำประจุด (♩.)เป็น 1 จังหวะเคาะ

งานวิจัยที่เกี่ยวข้อง

จากงานวิจัยของ Hendrik Schreiber และ Meinard Müller [1] กล่าวไว้ว่าวิธีการทั่วไปที่จะประมาณอัตราเร็วจังหวะ (tempo) จะทำโดยการหาจุดเริ่มต้นของเสียง (onset) และ จังหวะเคาะ (beat) จากนั้นจึงนำไปประมาณหาอัตราเร็วจังหวะ แต่สำหรับงานวิจัย [1] นี้จะประมาณอัตราเร็วจังหวะโดยตรงจากสเปกโตรแกรมของเมล (mel-spectrogram) แบบขั้นตอนเดียว โดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ (convolutional neural network) ซึ่งการประมาณค่าอัตราเร็วจังหวะนี้เป็นปัญหาการจำแนกประเภทแบบหลายหมวดหมู่ (multi-class classification) โดยการสอนโครงข่ายประสาทเทียมนั้นจะใช้ชุดข้อมูลเพลงที่ครอบคลุมแนวเพลงและอัตราเร็วจังหวะที่มีความหลากหลายควบคู่ไปกับการเพิ่มข้อมูล ผลการทดลองพบว่าการประมาณค่าอัตราเร็วจังหวะสามารถเอาชนะขั้นตอนวิธีที่ดีที่สุดในตอนนั้น แต่ยังไม่สามารถเอาชนะในกรณีที่มีความคลาดเคลื่อนของอัตราเร็วจังหวะคู่แปด (tempo octave error) ที่ความคลาดเคลื่อนยินยอมที่ 4% โดยความคลาดเคลื่อนของอัตราเร็วจังหวะคู่แปดคือปัญหาของผลลัพธ์ของขั้นตอนวิธีที่ได้อัตราเร็วจังหวะออกมาเป็นเศษส่วนหรือเป็นจำนวนที่เป็นทวีคูณ

จากงานวิจัยของ Jan Schlüter และ Sebastian Böck [2] ระบุไว้ว่าจุดเริ่มต้นของเสียงคือจุดเริ่มต้นที่มีการเกิดเสียงแต่ละเสียงในเพลง โดยการตรวจจับจุดเริ่มต้นของเสียง (onset detection) สามารถนำไปใช้ประโยชน์ในงานต่างๆได้ เช่น การตรวจจับจังหวะเคาะ การประมาณค่าอัตราเร็วจังหวะ และการวิเคราะห์เพลง เป็นต้น จากสเปกโตรแกรมของเพลงจะทราบว่าจุดเริ่มต้นของเสียงเป็นการเปลี่ยนแปลงของสเปกตรัมต่อเวลา เมื่อพิจารณาจากจุดนั้นทำให้ทราบว่าการหาจุดเริ่มต้นของเสียงนั้นมีความคล้ายคลึงกับการหาขอบในรูปภาพ จึงนำไปสู่การใช้โครงข่ายประสาทเทียมแบบสังวัตนาการในการหาจุดเริ่มต้นของเสียง ผลการทดลองได้ค่าความแม่นยำ (precision) 90.5% ค่าความครบถ้วน (recall) 86.6% และค่า F-measure 88.5%

คลังโปรแกรมที่เกี่ยวข้อง

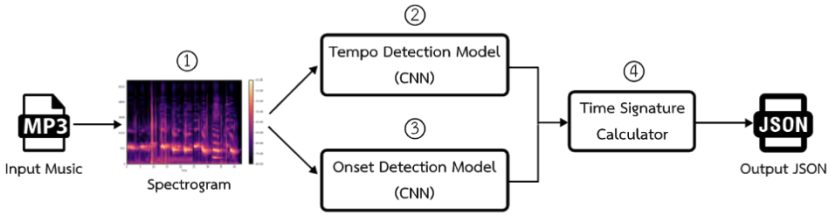
- Pydub [4] เป็นคลังโปรแกรมที่ช่วยในการจัดการกับแฟ้มข้อมูลเสียง ที่สามารถใช้ในการอ่านแฟ้มข้อมูลเสียง และสามารถตัดแต่ง และบันทึกเป็นแฟ้มข้อมูลใหม่
- LibROSA [5] เป็นคลังโปรแกรมที่มีขั้นตอนวิธีช่วยในการจัดการการวิเคราะห์ข้อมูลของเพลงและเสียง โดยบริการตรวจจับอัตราเร็วจังหวะนี้จะอาศัยการคัดแยกคุณลักษณะของเสียง การแปลงคลื่นเสียงให้กลายเป็นสเปกโตรแกรม และใช้การตรวจจับจุดเริ่มต้นของเสียง การตรวจจับจังหวะเคาะและอัตราเร็วจังหวะ เพื่อใช้ในการเรียนรู้ของผู้จัดทำ

ภาพรวมของระบบ

ภาพรวมของระบบเป็นดังแสดงในรูปที่ 3 แบ่งออกเป็น 4 ขั้นตอน

ขั้นตอนแรก เมื่อรับข้อมูลเป็นแฟ้มข้อมูลเอ็มพี 3 มาแล้ว จะแปลงให้เป็นสเปกโตรแกรมก่อนโดยใช้คลังโปรแกรม LibROSA จากนั้น ขั้นตอนที่สองจะระบุจังหวะเคาะด้วยตัวแบบตรวจจับจังหวะ (Tempo Detection Model) และขั้นตอนที่สามจะหาจุดเริ่มต้นของโน้ตด้วยตัวแบบตรวจจับจุดเริ่มต้นของโน้ต (Onset Detection Model) ในขั้นตอนสุดท้ายจึงนำข้อมูลจังหวะและจุดเริ่มต้นของโน้ตมาคำนวณหาเครื่องหมายกำหนดจังหวะด้วยโมดูลเครื่องคำนวณเครื่องหมายกำหนดจังหวะ (Time Signature Calculator)

ในการสร้างตัวแบบเพื่อตรวจจับอัตราเร็วจังหวะ และตรวจจับจุดเริ่มต้นของโน้ตนั้น จะใช้ข้อมูลเพลงจริงจากแฟ้มข้อมูลเอ็มพี 3 ของเพลงจากอัลบั้ม เปรียบเทียบกับข้อมูลของคุณลักษณะของเพลงจากเว็บ Spotify และใช้ประเมินผลความถูกต้องแม่นยำของตัวแบบที่สร้าง



รูปที่ 3

ขอบเขตของโครงงาน

- โครงงานนี้ศึกษารูปแบบของจังหวะเคาะ อัตราเร็วจังหวะ อัตราจังหวะ จุดเริ่มต้นของจังหวะและเครื่องหมายกำหนดจังหวะ เฉพาะดนตรีประเภท พอป ร็อกแอนด์โรล และ อาร์แอนด์บี เท่านั้น
- ข้อมูลเข้าและข้อมูลที่ใช้แฟ้มข้อมูลเอ็มพี 3 ที่มีความยาวเต็มเพลง
- ข้อมูลที่มีการกำกับผลลัพธ์สำหรับการฝึกสอนตัวแบบเป็นข้อมูลที่มาจาก Spotify WEB API
- ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการสำหรับการตรวจจับจังหวะและการตรวจจับจุดเริ่มต้นของโน้ต
- โครงงานนี้เป็นโครงงานพัฒนาบริการที่รับแฟ้มข้อมูลเสียงประเภทเอ็มพี 3 แล้วส่งออกผลลัพธ์เป็นรูปแบบ JSON ซึ่งจะระบุออกมาเป็นค่าอัตราเร็วจังหวะ (Tempo) และเครื่องหมายกำหนดจังหวะ (Time Signature) โดยมีรูปแบบดังนี้

  { “tempo” : 60,
  “Time-signature” : 4,}

วิธีการดำเนินงาน

1. ค้นหาและศึกษางานวิจัย โดยอ้างอิงขั้นตอนการทำงานจากงานวิจัยของ Hendrik Schreiber และ Meinard Müller [1] และ Jan Schlüter และ Sebastian Böck [2] ซึ่งเป็นวิธีการหาอัตราเร็วจังหวะและจุดเริ่มต้นของเสียง รวมถึงศึกษาเนื้อหาที่เกี่ยวข้องกับทฤษฎีดนตรีจากหนังสือทฤษฎีดนตรีสากลฉบับสรรพสูตร [6] จังหวะ (The Rhythm) [7] และทฤษฎีดนตรี [8]
2. ศึกษาเครื่องมือ โปรแกรมและเทคนิคที่ใช้ในงานวิจัย เพื่อเป็นประโยชน์ในการทำโครงงาน โดยศึกษาคลังโปรแกรม pydub [4] ศึกษาวิธีการจัดการกับแฟ้มข้อมูลเสียง และ LibROSA[5] ควบคู่กับงานวิจัยของ Brian McFee และคณะ [3] ศึกษาขั้นตอนกระบวนการทำงานของคลังโปรแกรม LibROSA
3. รวบรวมข้อมูล เพื่อกำหนดขอบเขตของโครงงาน
4. วิเคราะห์และออกแบบวิธีการ แบ่งการทำงานออกเป็น 4 ส่วนหลัก ๆ ได้แก่
    a. การจัดการข้อมูล
    b. การฝึกสอนตัวแบบการตรวจจับจังหวะ
    c. การฝึกสอนตัวแบบการตรวจจับจุดเริ่มต้นของโน้ต
    d. การคำนวณหาเครื่องหมายกำหนดจังหวะ
5. พัฒนาระบบตรวจจับอัตราเร็วจังหวะ
6. ตรวจสอบความถูกต้องและวัดประสิทธิภาพโดยตรวจสอบความถูกต้องจากชุดข้อมูลทดสอบ และค่าผิดพลาดกำลังสองเฉลี่ย (Mean Squared Error: MSE)
7. สรุปผลการดำเนินงาน
8. จัดทำเอกสาร

ตารางเวลาการดำเนินงาน

ขั้นตอนการดำเนินงาน	ปี 2562					ปี 2563			
	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.
1.ค้นหาและศึกษาบทความรวมถึงเนื้อหาที่เกี่ยวข้องกับจังหวะทางดนตรี									
2.ศึกษาเครื่องมือ โปรแกรมและเทคนิคที่ใช้ในงานวิจัย									
3.รวบรวมข้อมูล เพื่อกำหนดขอบเขตของโครงงาน									
4.วิเคราะห์และออกแบบวิธีการ									
5.พัฒนาบริการตรวจจับอัตราเร็วจังหวะ									
6.ตรวจสอบความถูกต้องและวัดประสิทธิภาพ									
7.สรุปผลการดำเนินงาน									
8.จัดทำเอกสาร									

ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ต่อผู้พัฒนา

1. มีความรู้ ความเข้าใจในด้านดนตรี
2. มีความรู้ ความเข้าใจในการทำบริการตรวจจับอัตราเร็วจังหวะ
3. เพิ่มพูนทักษะการเขียนโปรแกรมและการพัฒนาระบบ
4. เรียนรู้การคิดวิเคราะห์วางแผนการทำงานอย่างเป็นระบบแบบแผน เพื่อให้เกิดประโยชน์สูงสุดตามทรัพยากรที่มีอยู่
5. ฝึกการเรียนรู้ด้วยตนเอง การยอมรับฟังความคิดเห็นของผู้อื่น ความตรงต่อเวลา ความรับผิดชอบในหน้าที่

ประโยชน์ต่อผู้ใช้ระบบ

1. ผู้ใช้งานสามารถนำอัตราเร็วจังหวะที่วัดได้ไปใช้ประโยชน์ในด้านต่าง ๆ

2. ผลลัพธ์จากการตรวจจับอัตราเร็วจังหวะ และ เครื่องหมายกำหนดจังหวะมีความถูกต้องและแม่นยำ

อุปกรณ์และเครื่องมือที่ใช้

1. ฮาร์ดแวร์
    1. แล็ปท็อปคอมพิวเตอร์ รุ่น Intel® Core™ i5-8210Y ความเร็ว 1.60 GHz RAM 8.00 GB
    2. แล็ปท็อปคอมพิวเตอร์ รุ่น Intel® Core™ i7-5500U ความเร็ว 2.40 GHz RAM 8.00 GB
    3. AWS Amazon SageMaker 4 vCPU Memory 64 GiB 1xK80 GPU GPUMemory 12 GiB
2. ซอฟต์แวร์
    1. Google Chrome
    2. Microsoft Office 365
    3. Jupyter Notebook
    4. Visua Studio Code

งบประมาณ

1. กระดาษถ่ายเอกสารขนาด A4 80 แกรม	ราคา	300	บาท
2. หมึกพิมพ์	ราคา	1500	บาท
3. ค่าถ่ายเอกสารและค่าทำรูปเล่ม	ราคา	300	บาท
4. ค่าเช่า AWS Amazon SageMaker	ราคา	5000	บาท
5. ค่าอุปกรณ์จัดเก็บบันทึกข้อมูล	ราคา	1500	บาท
6. ค่าอุปกรณ์อิเล็กทรอนิกส์	ราคา	10000	บาท
7. ค่าบำรุงรักษาอุปกรณ์	ราคา	1500	บาท

20100 บาท

<u>หมายเหตุ</u> ทั้งนี้งบประมาณอาจเปลี่ยนแปลงตามความเหมาะสมและขอถัวเฉลี่ยทุกรายการ

เอกสารอ้างอิง

[1] Hendrik Schreiber, Meinard Müller. A Single-Step Approach To Musical Tempo Estimation Using A Convolutional Neural Network. Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR) (2561) : 98-105

[2] Jan Schlüter, Sebastian Böck. Musical Onset Detection with Convolutional Neural Networks. 6th International Workshop on Machine Learning and Music (MML) (2556)

[3] Brian McFee, Colin Raffel, Dawen Liang, Daniel P.W. Ellis, Matt McVicar, Eric Battenbergk, Oriol Nieto. LibROSA: Audio and Music Signal Analysis in Python. PROC. OF THE 14th PYTHON IN SCIENCE CON. SCIPY (2558)

[4] James Robert. Pydub.[ระบบออนไลน์]. แหล่งที่มา : https://github.com/jiaaro/pydub [1 กันยายน 2562]

[5] Brian McFee, Stefan Balke. LibROSA.[ระบบออนไลน์] แหล่งที่มา : https://LibROSA.github.io/LibROSA/ [1 กันยายน 2562]

[6] สำเร็จ คำโมง. ทฤษฎีดนตรีสากลฉบับสรรพสูตร. กรุงเทพมหานคร : สำนักพิมพ์ฐานบัณฑิต. 2553

[7] สมศักดิ์ สร้อยระย้า. จังหวะ (The Rhythm). กรุงเทพมหานคร : สำนักพิมพ์โอเดียนสโตร, 2545.

[8] ณัชชา พันธุ์เจริญ. ทฤษฎีดนตรี. พิมพ์ครั้งที่ 13. กรุงเทพมหานคร : สำนักพิมพ์เกศกะรัต, 2558.

[9] ศานติ เดชคำรณ. เอกสารประกอบการสอน ทฤษฎีดนตรีสากล 1. เอกสารประกอบการสอน ทฤษฎีดนตรีสากล 1 คณะศิลปกรรมศาสตร์ : มหาวิทยาลัยบูรพา, 2548.

ประวัติผู้จัดทำ

Mr. Patchara Charoenkij

นาย พชระ เจริญกิจ

วันเกิด 18 กันยายน 2540

โรงเรียนตราษตระการคุณ

ชั้นปีที่ 4 คณะวิทยาศาสตร์

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

สาขาวิชาวิทยาการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย

อีเมล: p.charoenkij@gmail.com

Mr. Surakiat Robkhob

นาย สุรเกียรติ รอบคอบ

วันเกิด 30 เมษายน พ.ศ. 2541

โรงเรียนราชวินิตบางแก้ว

ชั้นปีที่ 4 คณะวิทยาศาสตร์

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

สาขาวิชาวิทยาการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย

อีเมล: surakiat.drome@gmail.com