

การจำแนกเพลงไทยและเพลงฝรั่งด้วยโครงข่ายประสาทเทียมโดยใช้ คุณลักษณะเชิงคุณภาพเสียง

Thai English Music Classification by Artificial Neural Network Based on Timbral Features

นายลักษณะันท์ แก้วทงศ์และ ณัฐธนนท์ หงส์วิทธิธร

บทคัดย่อ

การจัดหมวดหมู่ประเภทเพลงสำหรับการจัดการฐานข้อมูลเสียงเพลงนั้นจะเป็นประโยชน์ต่อการสืบค้นฐานข้อมูลเพลงที่เพิ่มขึ้นอย่างต่อเนื่องในแต่ละปีทั้งเพลงไทยและเพลงฝรั่ง จึงจำเป็นต้องใช้บุคคลากรที่มีความชำนาญอย่างยิ่ง การหาระบบช่วยทุ่นแรงและลดความผิดพลาดในการจำแนกประเภทเพลงที่มีประสิทธิภาพจึงเข้ามามีบทบาทต่อการจำแนกประเภทเพลงที่ใช้ภาษาที่ต่างกันจึงเป็นสิ่งที่น่าสนใจ

ในงานวิจัยชุดนี้นำเสนอการจำแนกเพลงไทยและเพลงฝรั่งด้วยโครงข่ายประสาทเทียมแบบหลายชั้นโดยใช้ค่าคุณลักษณะเชิงคุณภาพเสียง (Timbral Feature) ในการทดลองส่วนแรกการจำแนกเพลง 8 ประเภท 1 ขั้นตอนจะปรับค่าพารามิเตอร์ที่สำคัญ 5 ค่าคือ Hidden Unit, Learning Rate, Training Time, Momentum และ Threshold โครงสร้าง (Topology) จำนวนโหนด Input เท่ากับ 34 จำนวนโหนดของ Hidden Unit เท่ากับ 30 และ จำนวนโหนดของ Output เท่ากับ 8 ใช้วิธีทดสอบแบบการทดลองแบบทวนสอบ 10-Fold Cross Validation ในส่วนของขั้นตอนเป็นการปรับค่าพารามิเตอร์ (Parameter Tuning) ของโครงข่ายประสาทเทียมแบบหลายชั้นโดยจะทำการปรับค่าพารามิเตอร์เพื่อที่จะหาค่าที่ดีที่สุด ในการทดลองนี้ค่าพารามิเตอร์ที่ดีที่สุดคือ Learning Rate เท่ากับ 0.75 Momentum เท่ากับ 0.25 Hidden Unit เท่ากับ 30 Training Rate เท่ากับ 3000 และ Threshold เท่ากับ 40 ผลการวิเคราะห์ค่ากลางพบว่าจากการทดลองใช้ค่าทางสถิติ 7 กลุ่มเพื่อเป็นค่ากลางพบว่าค่าที่นำมาใช้ในการจำแนกแล้วได้ค่าในการจำแนกที่ดีที่สุดคือ ค่า Mean - Median ผลการเปรียบเทียบกลุ่มค่าคุณลักษณะที่ดีที่สุดในช่วงกึ่งกลางเพลงที่ 10 วินาที ZC+SR+MFCC+LPC+SC+SF ที่ 74.45% ค่ากลุ่มคุณลักษณะการจำแนก 10 วินาทีแรกของเพลง MFCC+SR+LPC ที่ 67.32% ค่ากลุ่มคุณลักษณะการจำแนก 20 วินาทีแรกของเพลง ZC+MFCC+SSB+LPC+SC+SF ที่ 58.72% ค่า

กลุ่มคุณลักษณะการจำแนก 30 วินาทีแรกของเพลง ZC+LPC+MFCC+SF+SC+SSB ที่ 58.72% ผลการจำแนกโดยใช้ขนาดเฟรมจากการทดลองครั้งนี้เลือกใช้ขนาดเฟรม 5 ค่า 128 256 512 1024 และ 2048 พบว่าค่าที่ขนาดเฟรมที่ใช้ในการจำแนกได้ค่าดีที่สุด 512 หลังจากนั้นหาค่าที่ดีที่สุดในกลุ่มของ MFCC 0 - 4 พบว่าค่า MFCC 1 ให้ค่าดีที่สุด หลังจากนั้นนำค่า MFCC 1 ไปเปรียบเทียบกับค่าคุณลักษณะที่มีค่าเดียวพบว่า MFCC 1 เป็นรองค่า SR ค่าเดียว ในส่วนที่ 2 ส่วนการจำแนกเพลงแบบ 8 ประเภท 2 ขั้นตอนพบว่าค่าในขั้นตอนแรกการจำแนกเพลงที่ใช้ภาษาไทยและฝรั่งตั้งแต่กลุ่มที่มี 2 ค่าคุณลักษณะขึ้นไปสามารถจำแนกประเภทเพลงได้ดีที่ 99.75% ขั้นตอนที่สองการจำแนกประเภทเพลงที่ใช้ภาษาไทย กลุ่มค่าคุณลักษณะที่ได้ค่าในการจำแนกดีที่สุดคือ ZC+SR+MFCC+LPC+SC+SF ให้ค่าความถูกต้องที่ 72.47% ส่วนการจำแนกประเภทเพลงที่ใช้ภาษาฝรั่งกลุ่มค่าคุณลักษณะที่ได้ค่าในการจำแนกดีที่สุดคือ ZC+SR+MFCC+LPC ให้ค่าความถูกต้องที่ 73.43% ในส่วนสุดท้ายเพลงที่ใช้ทำนองเดียวกันแต่ใช้ภาษาแตกต่างกันพบว่าค่า ZC ให้ค่าในการจำแนกสูงสุดที่ 61.54%

Abstract

Effective classification of music in the management of music databases is very important for the user search activity on Thai – English music databases is increasing continuously year by year. Because of the increased volume and complexity it is necessary to use skilled personnel to administer the databases. Therefore, finding an efficient system which helps reduce mistakes in music classification plays an important role in make databases more user - friendly, especially when dealing with songs in different languages.

This research presents a classification of Thai – English songs with a multi – layer artificial neural network using timbral features. The first part of the experiment related to the classification of eight song categories. The researcher adjusted five important parameters, namely, hidden unit, learning rate, training time, momentum, and threshold.

The topology of input comprised 34 nodes; hidden unit comprised 30 nodes; and output unit comprised eight nodes. The test was conducted with 10 – fold cross validation. To find the best tuning parameter, the researcher tuned the parameter of the multi – layer artificial neural network to. The optimal parameters of identified by this experiment were as follows: Learning Rate, 0.75; Momentum, 0.25; Hidden Unit, 30; Training Time, 3000; and Threshold, 40. According to the result of analyzing the middle value; after using seven groups of statistical values as the middle value, the researcher found that the best value used in classification was the Mean – Median. According to comparison results, the best feature groups which could classify a song at 10 seconds were ZC+SR+MFCC+LPC+SC+SF at 74.45%. MFCC+SR+LPC were the feature which could classify a song in the first 10 second at 67.32%. ZC+MFCC+SSB+LPC+SC+SF were the feature which could classify the song in the first 20 second at 58.72%. ZC+LPC+MFCC+SF+SC+SSB were the feature which could classify a song in the first 30 second at 58.72%. The result were also classified using five values of frame; 128, 256, 512, 1024, and 2048. The researcher found that the value of frame best used in classification was 512. After finding the best value of MFCC 0 - 4, the researcher found the MFCC 1 gave the best value. After comparing the value of MFCC 1 with had a value, the researcher found that MFCC 1 was less than the SR value for a value only.

Part 2 of the research the classification of eight song categories in two steps. The researcher found that the values in the first described above, relate to classification of Thai – English songs form a group with more than two features, could classify song categories with an accuracy of 99.75%.The second step related to classification of Thai song. Part 3 of the research found that the feature which could best classify were

ZC+SR+MFCC+LPC+SC+SF giving an accuracy level of ZC+SR+MFCC+LPC giving an accuracy level of 73.43%.As for the songs using the same melody but different languages, the researcher found that ZC value was most effective at 61.54%.