

บทที่ 2

งานวิจัยที่เกี่ยวข้อง

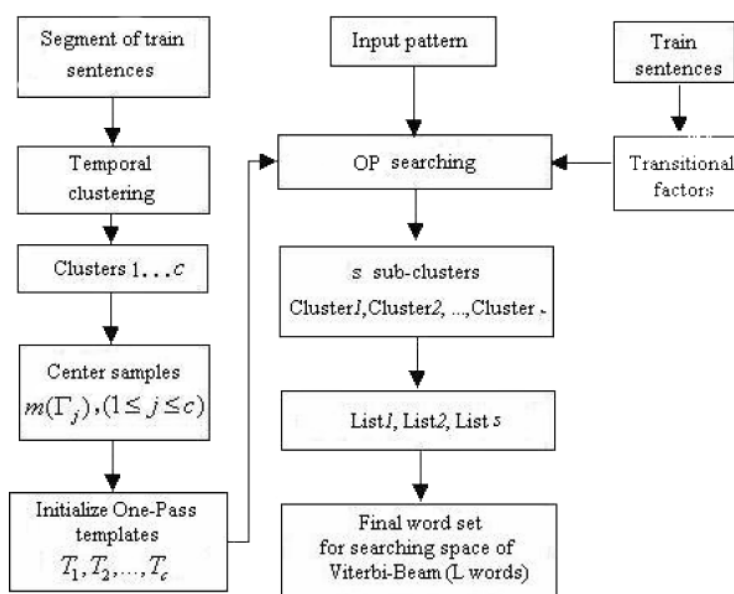
งานวิจัยเกี่ยวกับการแปลภาษามือมีหลายๆ งานวิจัยที่ผ่านมา ดังนี้

Gaolin Fang และ Wen Gao [1] ได้ใช้อุปกรณ์ถุงมืออิเล็กทรอนิกส์ร่วมกับอุปกรณ์ตรวจจับตำแหน่ง 6 แกน ได้มีการรู้จำท่ามือต่อเนื่องที่เป็นแบบ Signer-Independent วิธีที่ใช้คือได้ทำการ pre-segment ด้วย Simple Recurrent Network (SRN) เพื่อที่จะตัดการเคลื่อนไหวระหว่างท่ามือ (Transition Movement) ออกก่อนที่จะนำข้อมูลมารู้จำด้วยฮิดเดนมาร์คอฟโมเดล ทำให้ไม่ต้องสร้างโมเดลของการเคลื่อนไหวระหว่างท่ามือ เพื่อที่จะทำให้ sequence สั้นลง ทำให้เวลาที่ใช้ Viterbi algorithm ในการค้นลำดับของคำที่ดีที่สุดก็จะสั้นลงด้วย ในการสร้างโมเดลนั้นสเตทของฮิดเดนมาร์คอฟโมเดลอยู่ระหว่าง 1-4 สเตท ซึ่งขึ้นอยู่กับความยาวของข้อมูล ในการทดลองมีการเก็บท่ามือจากผู้ทำท่ามือทั้งหมดจำนวน 7 คน ซึ่งกระทำทำหนึ่ง 208 ท่าจำนวนท่าละ 3 ครั้ง สำหรับประโยคต่อเนื่องจำนวน 100 ประโยคจะประกอบไปด้วยท่า 208 ท่าดังกล่าว จากผลการทดลองนั้นทำให้ทราบได้ว่ารู้จำท่ามือจากผู้ใช้ 5 ใน 7 คน มีค่าเฉลี่ย 95.3% ปัญหาคือ ถ้า segment ผิดจะมีผลต่อเนื่องกันไป ซึ่งทำให้ความถูกต้องต่ำกว่าการใช้ฮิดเดนมาร์คอฟโมเดลเพียงอย่างเดียว

Wen Gao, Gaolin Fang, Debin Zhao และ Yiqiang Chen [2][3] ได้นำเสนอการพัฒนาวิธีการแก้ปัญหาคำศัพท์ที่มีจำนวนมากมายของท่าต่อเนื่อง ซึ่งเน้นการสร้างโมเดลของการเคลื่อนไหวระหว่างท่ามือด้วยการ Segment ในส่วนของการเคลื่อนไหวระหว่างท่ามือก่อน แล้วนำมา Clustering การเคลื่อนไหวระหว่างท่ามือด้วย Dynamic Time Warping (DTW) เมื่อได้ข้อมูลที่ Cluster แล้วก็นำมา train ให้แก่ระบบ แล้วใช้ Transition Models ที่ได้มาใช้ในการ segment ประโยคที่ต่อเนื่องให้เป็นท่าย่อยๆ ซึ่งได้ทำการทดลอง 25,565 ท่าจากท่าเดียว 5,113 ท่า โดยแต่ละท่าท่าละ 5 ครั้ง และเลือกข้อมูล 4 ชุดมาใช้ในการกระบวนการสอนและอีกหนึ่งชุดที่เหลือใช้สำหรับการทดสอบ ซึ่งผลการทดลองมีความถูกต้อง 95.4% สำหรับท่าเดียว 5,113 ท่า และสำหรับท่าระหว่างมือจาก 1,500 ประโยค (เก็บข้อมูล 750 ประโยคเป็นจำนวน 2 ครั้ง) โดยข้อมูลชุดแรกใช้ในการกระบวนการสอน และอีกชุดใช้สำหรับการทดสอบ ในหนึ่งประโยคมีค่าโดยเฉลี่ยคือ 6.6 คำ ซึ่งผลที่ได้คือมีความถูกต้อง 90.8% ซึ่งวิธีการนี้จะพบปัญหาคือ การ train ท่าเคลื่อนไหวระหว่างมือต้องใช้เวลา และต้อง train ให้ครอบคลุมจึงจะนำไปทำการ Cluster ได้

Jianjun Ye, Hongxun Yao และ Feng Jiang [4] ได้ใช้ฮิดเดนมาร์คอฟโมเดลร่วมกับ SVM (Support Vector Machines) รวมเป็น 2 ส่วนในการรู้จำบน Confusable set และได้ใช้ฮิดเดนมาร์คอฟโมเดลในการรู้จำท่าปกติ ซึ่งจะเป็นตัวกำหนดขอบเขตของท่าที่ไม่สามารถแยกได้ (Confusability) ในตอนแรก และใช้ SVM ในการแก้ปัญหาความไม่แน่นอนหลังจากที่ใช้ฮิดเดน

มาร์คอฟโมเดลในการรู้จำ ซึ่งในงานวิจัยนี้ได้พบว่าหากมีทำเป็นจำนวนมากๆ ก็จะมีบางทำที่มีความคล้ายคลึงกัน ซึ่งจะทำให้การแบ่งแยกมีคุณภาพลดลง จึงนำฮิดเดนมาร์คอฟโมเดลใช้ร่วมกับ SVM เพื่อให้การรู้จำมีประสิทธิภาพมากขึ้น หากมีข้อมูลของทำมือเข้ามาก็จะทำการรู้จำด้วยฮิดเดนมาร์คอฟโมเดล ถ้าข้อมูลไม่ได้อยู่ในกลุ่มของ Confusable set ก็จะพิจารณาว่าเป็นผลของการรู้จำได้เลย แต่ถ้าอยู่ในกลุ่มของ Confusable set ก็จะรู้จำทำมือจาก Confusable sign ใน Confusable set ซึ่งได้ทำการทดลองกับ 4,942 ทำ ซึ่งกระทำโดยผู้ทำทำจำนวน 6 คน (รวมเป็น 59,304 ทำ) โดยแบ่งข้อมูลจากผู้ทำทำคนละ 2 ชุด ซึ่งจาก 12 ชุด จะใช้ 1 ชุดสำหรับการทดสอบ และ 11 ชุดสำหรับกระบวนการสอน ซึ่งได้ผลการรู้จำ 89.40% ซึ่งในงานวิจัยนี้ยังไม่ได้นำมาใช้กับทำมือที่ต่อเนื่อง



รูปที่ 2.1 ไคอะแกรมของ OP pre-searching

Guilin Yao, Hongxun Yao, Xin Liu และ Feng Jiang [5] ได้นำเสนอถึงความเร็วในการรู้จำสำหรับจำนวนข้อมูลที่มีมากมาย ซึ่งจะมีการเฉลี่ยของเวลาเป็นระยะเวลานาน ดังนั้นจึงได้ใช้วิธีการ One-Pass (OP) ในการ pre-searching (แสดงดังรูปที่ 2.1) เพื่อค้นหาแต่ละ Sub-Cluster ที่เป็นไปได้เท่านั้น คือแบ่งประโยคให้เป็นส่วนย่อยด้วยการ Cluster ทำมือที่มีความคล้ายคลึงกันนั่นเอง ซึ่งได้ทำการหาถึงกลางของคำระหว่าง 2 คำแล้วทำการ cluster จะพบว่าความเร็วในการรู้จำและความถูกต้องนั้นมีความสัมพันธ์กันกับจำนวน Cluster ที่ได้จาก Temporal Clustering Algorithm และซึ่งการทดลองได้กระทำกับฐานข้อมูลขนาดใหญ่ โดยใช้ทำเดียวจำนวน 4,942 ทำ โดยแต่ละทำทำจำนวน 4 ครั้ง จากผู้ทำทำ 2 คน ซึ่งจะใช้ข้อมูลทั้งหมดในกระบวนการสอน และทำต่อเนื่องจำนวน 543 ประโยค โดยกระทำประโยคละ 4 ครั้งเช่นกัน แล้วทำการเลือกข้อมูล 3 ชุดสำหรับการแบ่งกลุ่มและการ segment ด้วย Viterbi-beam และอีก 1 ชุดใช้ในการทดสอบ โดยแต่ละประโยค

ประกอบด้วยคำโดยเฉลี่ย 6.10 คำ ซึ่งผลการทดลองถ้ายิ่งแบ่งกลุ่มมากขึ้น ความเร็วในการรู้จำจะมีมากขึ้นแต่ความถูกต้องจะลดลงงานวิจัยนี้ได้พบว่ามีข้อจำกัดคือ ในการใช้ Center matching method บางการรู้จำจะผิดพลาดได้ และต้องไม่มีการเคลื่อนไหวระหว่างท่ามือระหว่าง Center และจากการใช้ Transition factor ในการ Classify ของ OP-searching จะมีข้อดีคือ ทำให้การรู้จำแบบ real time ดีกว่าการใช้ Viterbi-beam algorithm เพียงอย่างเดียว เนื่องจาก error อาจเกิดขึ้นในขั้นตอนแรกคือ การ Classify ผิด Cluster

วุฒิชัย วิศาลคุณา [14] ได้นำเสนอการจดจำภาษามือไทยโดยใช้ฮิดเดนมาร์คอฟโมเดลโดยใช้ถุงมืออิเล็กทรอนิกส์ร่วมกับอุปกรณ์ตรวจวัดตำแหน่ง 6 แกนเป็นตัวรับข้อมูล แล้วนำข้อมูลไปผ่านขั้นตอนประมวลผลเบื้องต้น เพื่อทำการแบ่งข้อมูลออกเป็นชุด คือ รูปแบบของมือ ตำแหน่งของมือ การหมุนของมือ และการเคลื่อนที่ของมือ เพื่อทำการหาจุดเริ่มต้นและจุดสิ้นสุดของภาษามือท่าหนึ่ง ซึ่งจะเป็นการแบ่งแยกภาษามือที่เป็นท่าหนึ่งและท่าเคลื่อนไหว ซึ่งจะทำให้สามารถแบ่งแยกคำในการทำภาษามือที่ทำต่อเนื่องกันได้ แต่หากหาขอบเขตของภาษามือไม่ดี จะทำให้โมเดลภาษามือไทยที่ประมาณได้มีคุณภาพไม่ดี ในส่วนของฮิดเดนมาร์คอฟโมเดลนั้น ได้นำข้อมูลจากขั้นตอนการประมวลผลเบื้องต้นมาทำการสอนให้ระบบแล้วจึงทำการจดจำข้อมูลภาษามือชุดนั้น หลังจากทำการสอนให้ระบบจดจำภาษามือแล้ว เมื่อผู้ทำภาษามือลักษณะเดียวกับแบบที่ทำการสอนระบบจะสามารถบอกได้ว่าผู้ทำภาษามือคำใด หลังจากหาภาษามือที่เหมาะสมกับข้อมูลในขณะเวลาดังกล่าวได้แล้ว จึงนำข้อมูลที่ได้อามาเรียงกันเป็นประโยคที่ต้องการตามหลักไวยากรณ์และความหมาย โดยใช้หลักการทางการประมวลผลภาษามือธรรมชาติ ซึ่งงานวิจัยนี้ไม่ได้วิจัยในส่วนของการเคลื่อนไหวระหว่างท่ามือและเป็นแบบ Signer dependence

วรวิทย์ วีระพันธุ์ [6] ได้ทำการศึกษาท่ามือภาษาไทย ซึ่งท่ามือภาษาไทยนั้นแบ่งออกเป็น 2 ส่วน คือ ส่วนที่เป็นท่ามือนิ่ง (Static Gesture or Posture) และส่วนที่เป็นท่ามือที่มีการเคลื่อนไหวของมือ (Dynamic Gesture) ซึ่งปัญหาที่พบในการรู้จำท่ามือที่เป็นท่าต่อเนื่อง (Continuous Gesture) ก็คือ จะมีวิธีการอย่างไรที่จะสามารถแยกการเคลื่อนไหวของท่ามือ ซึ่งจะประกอบไปด้วยการเคลื่อนไหวของมือที่มีความหมาย (Meaningful Dynamic Gesture) และการเคลื่อนไหวของมือที่ไม่มีความหมายในตัวเอง (Transitional Gesture) นอกจากนั้นยังมีปัญหาเรื่องความเร็วในการทำท่ามือ, ตำแหน่งของท่ามือ และทิศทางการหันฝ่ามือ ซึ่งแต่ละคนจะมีการทำที่ไม่เหมือนกันหรือไม่เท่ากัน โดยในงานวิจัยนี้ยังได้นำเสนอวิธีการในการแก้ไขปัญหาเหล่านี้ โดยยึดหลักที่ว่าท่ามือภาษาไทยและท่ามือภาษาอื่น การเคลื่อนไหวของมือที่มีความหมายนั้น จะเป็นการเคลื่อนที่แบบซ้ำๆ (Periodic) ดังนั้นการแยกท่ามือที่มีการเคลื่อนไหวของมือ (Dynamic Gesture) ออกจากการเคลื่อนที่ของมือที่ไม่มีความหมายได้ โดยใช้ Fourier Analysis กับข้อมูลในแต่ละช่วงเวลา ซึ่งข้อมูลที่ได้โดยการใช้ถุงมืออิเล็กทรอนิกส์ที่เรียกว่า CyberGlove ในกรณีของท่ามือที่มีการเคลื่อนไหวของมือ (Dynamic Gesture) การประมาณค่าช่วงสัญญาณของท่ามือที่ได้จากการวิเคราะห์สามารถนำไปทำ Time Normalization ดังนั้น จึง

สามารถลดความผิดพลาดเนื่องจากการทำท่ามือที่มีความเร็วไม่เท่ากันของแต่ละคนได้ การทดลองจะใช้ข้อมูลที่เป็นท่ามือนิ่งและท่ามือที่มีการเคลื่อนไหวอย่างละ 100 ท่า และใช้วิธีการรู้จำที่เรียกว่า Bayesian Estimator ซึ่งผลการทดลองที่ได้ปรากฏว่ามีความถูกต้องในการรู้จำถึง 100% แต่วิธีการในงานวิจัยนี้ยังไม่สามารถรู้จำท่าเคลื่อนไหวแบบไม่ซ้ำได้