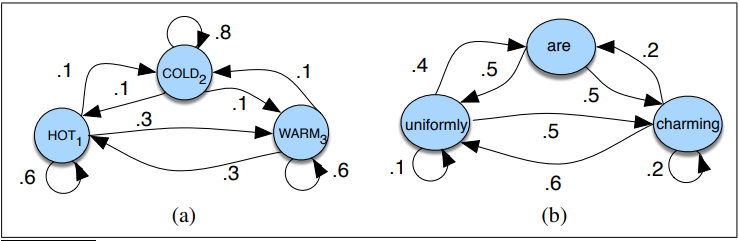
**บทที่ 2**

**ทฤษฎีและงานวิจัยที่เกี่ยวข้อง**

ในการศึกษาหาโมเดลที่จะมาคอยทำนายเส้นทางนั้นอาจจะประกอบไปด้วยหลายๆเทคนิคหรือหลายๆความรู้เข้าด้วยกัน ดังนั้นในบทนี้จะเป็นบทที่รวบรวมทฤษฎีและงานวิจัยที่เกี่ยวข้องและอธิบายให้เข้าใจง่าย ๆ

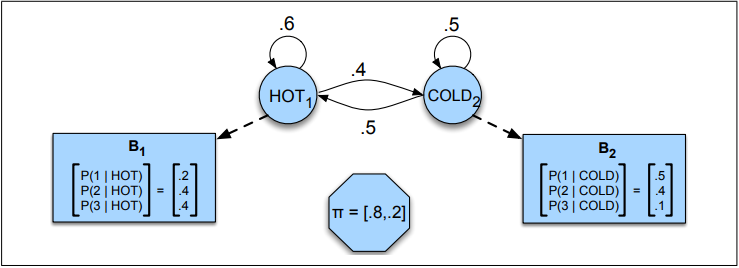
1. **ห่วงโซ่มาร์คอฟ (Markov Chains)**

คือโมเดลที่บอกบางอย่างเกี่ยวกับความน่าจะเป็นของลำดับตัวแปรสุ่ม ซึ่งเรียกว่า state โดยตัวแปรสุ่มนี้จะเป็นค่าอะไรก็ได้ตามที่เราสนใจ อาจจะเป็น คำ หรือสัญลักษณ์ที่แสดงถึงสิ่งต่าง ๆ เช่น สภาพอากาศ เป็นต้น ห่วงโซ่มาร์คอฟจะสร้างสมมุติฐานที่แม่นยำมากถ้าเราจะทำนายอนาคตของ สถานะ(state) ต่าง ๆที่เป็นลำดับ ซึ่ง สถานะ(state) ก่อนหน้าจะไม่ส่งผลต่อ สถานะ(state) ในอนาคตที่จะทำนายยกเว้น สถานะ(state) ปัจจุบัน เช่น ถ้าเราอยากจะทำนายสภาพอากาศของพรุ่งนี้ เราจะดูสภาพอากาศของวันนี้เพื่อทำนายสภาพอากาศวันถัดไป โดยจะไม่ดูสภาพอากาศของเมื่อวาน [1]

รูปที่ 3.1 Markov Chains

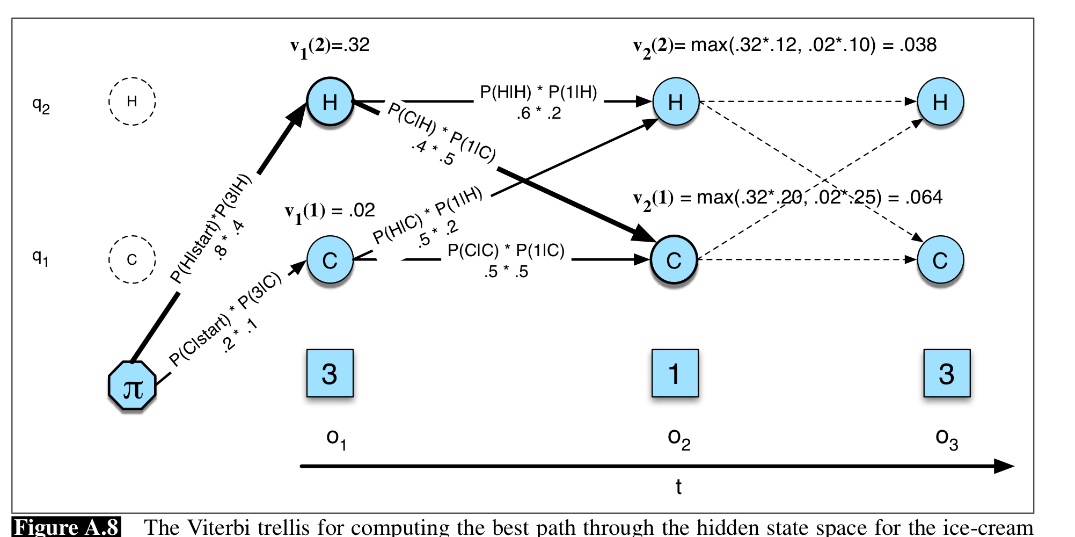
1. **แบบจำลองมาร์คอฟซ่อนเร้น (Hidden Markov Chains)**

คือโมเดลที่มีรากฐานมาจาก ห่วงโซ่มาร์คอฟ (Markov Chains) ซึ่งจะมีประโยชน์ก็ต่อเมื่อเรารู้ความน่าจะเป็นของลำดับ สถานะ(state) ที่เราสนใจ ถ้า สถานะ(state) ที่เราสนใจนั้นซ่อนอยู่ โดยเราจะไม่สังเกต สถานะ(state) โดยตรง เช่น เราจะทำนายสภาพอากาศวันนี้ โดยจะดูจากจำนวนไอศกรีมที่เพื่อนกินในวันนี้ ถ้ากิน 3 แท่ง เราจะทำนายว่าสภาพอากาศวันนี้นั้นร้อน โดยแบบจำลองมาร์คอฟซ่อนเร้น (Hidden Markov Chain)จะพูดถึง สถานะซ้อนเร้น (Hidden state)หรือก็คือสภาพอากาศวันนี้ที่เราจะทำนาย และตัวสังเกต(Observations) หรือก็คือจำนวนไอศกรีมที่เพื่อนกินวันนี้นั่นเอง [2]

****

รูปที่ 3.2 HIdden Markov Model

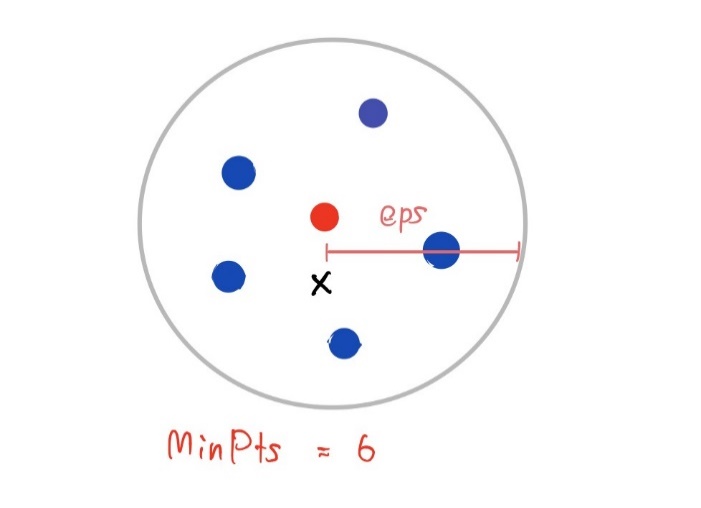
1. **Viterbi**

เป็น Algorithm ที่ใช้สำหรับหาลำดับ Hidden state ที่มีความน่าจะเป็นมากที่สุด จากการดู Observation ใน Hidden Markov Model [3]

รูปที่ 3.3 Viterbi Algorithm

1. **DBSCAN**

เป็นการ Clustering หรือการจัดกลุ่มของข้อมูล เป็นหนึ่งใน method ของ Unsupervised Learning โดย DBSCAN เป็นการหาบริเวณที่ข้อมูลเกาะกลุ่มกัน ซึ่งสามารถคำนวณได้จาก data point ที่อยู่รอบๆ ในรัศมีที่กำหนด โดยจะใช้พารามิเตอร์สองตัว คือ esp เป็นรัศมีจากจุดกลาง และ MinPts คือจำนวน Data Point ขั้นต่ำสำหรับการกำหนด Center[4]



รูปที่ 3.4 DBSCAN

1. **HDBSCAN**

เป็น Algorithm ที่ต่อยอดมาจาก DBSCAN ที่ใช้การทำงานของ Hierarchical Clustering มาร่วมด้วย ทำให้สามารถกลุ่มข้อมูลที่มีความหานแน่นต่างกันได้(มีค่า eps ได้หลายค่า) โดยใช้แค่ Parameter min\_cluster บอกจำนวนกลุ่มขั้นต่ำเท่านั้น โดย HDBSCAN สามารถแก้ปัญหาของ DBSCAN ในการใช้ esp ค่าเดียวในการจัดกลุ่มข้อมูลทุกกลุ่ม ทำให้บางกลุ่มอาจดูเหมือนหลายๆกลุ่มที่มาเกาะกัน

Chart, scatter chart

Description automatically generated

รูปที่ 5 ภาพเปรียบเทียบ DBSCAN (ซ้าย) และ HDBSCAN(ขวา)

1. **Trip destination prediction based on past GPS log using a Hidden Markov Model**

เป็นงานวิจัยที่เกี่ยวกับการคาดเดาเส้นทางการขับรถ โดยใช้ Model Hidden Markov Model โดยนำข้อมูลประวัติ GPS ของการเดินทางต่าง ๆ มาหาจุด Support Point ซึ่งเป็นจุดที่นำเส้นทางการขับรถมาหาจุดตัดที่รวมกันของทางแยก เพื่อลดปัญหาการคำนวณเนื่องจากข้อมูลประวัติ GPS นั้นมีจำนวนมาก เมื่อได้จุด Support Point แล้ว ก็จะนำมาเข้า Model Hidden Markov Chains โดยให้ Support Point เป็นตัว Observation ที่เราจะสังเกตแล้วนำมาคาดเดาเส้นทางและจุดหมายปลายทางของการเดินทางนั้นหรือก็คือ Hidden State นั้นเอง [6]

1. **Bayesian classifier for Route prediction with Markov chains**

เป็นงานวิจัยที่เกี่ยวกับการนำ Pattern การเดินทางของรถแต่ละคัน มาทำนายจุดหมายปลายทาง โดยถ้า pattern ที่มีจุด หมายเดียวกันจะถูกจัดกลุ่มเป็น Cluster เดียวกัน ซึ่งจะใช้ Markov Chains ในการTrain ว่าทริปการเดินทางไหนจะอยู่ Cluster ไหนและใช้ Bayesian Update ในการทำนายระหว่างทางว่าเส้นทางที่เลือกเดินเป็นของ Cluster ใด ก็จะสามารถรู้จุดหมายโดยให้หยุดทำนายเมื่อเปอร์เซ็นต์การทำนาย มากกว่าค่าค่าหนึ่ง เช่น 90% [7]

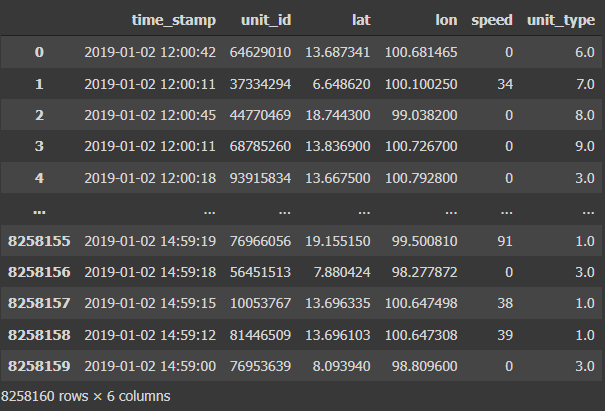
1. **TRIP/STOP Detection in GPS Traces to Feed Prompted Recall Survey**

เป็นงานวิจัยที่เปรียบเทียบการหาข้อมูลรถวิ่งจากข้อมูล GPS data โดยการใช้ Algorithm STOP detection และ TRIP detection โดย STOP detection คือการตรวจจับข้อมูลรถที่หยุดเคลื่อนที่ตามที่เรากำหนดว่าต้องไม่เคลื่อนที่เป็นเวลาเท่าใด ส่วน TRIP detection คือการตรวจจับข้อมูลรถที่เคลื่อนที่ [8]

**บทที่ 3**

**การออกแบบและการพัฒนา**

1. **Data set ที่ใช้**
2. ข้อมูล GPS ของปี 2019 โดยมีรายละเอียดดังนี้



|  |  |  |
| --- | --- | --- |
| Field | Data Type | Description |
| time\_stamp | Date time | วันเวลา |
| unit\_id | String | ID ของรถ |
| lat | double | latitude |
| lon | double | longitude |
| speed | float | ความเร็วในหน่วย km/h |
| unit\_type | int | ชนิดของรถ  1 = รถโดยสารประจำทาง  3 = รถโดยสารไม่ประจำทาง  4 = รถโดยสารส่วนบุคคล  5,8,9 = รถบรรทุกส่วนบุคคล  6,7 = รถบรรทุกไม่ประจำทาง |

1. **Data processing**
   1. การคัดกรองข้อมูลจีพีเอส

โดยเราจะนำข้อมูลจีพีเอสมาคัดกรองเอาเฉพาะพื้นที่ในภาคอีสานเท่านั้นโดยจะใช้ข้อมูล จีโอเมททรี ของจังหวัดในภาคอีสานเพื่อเช็คว่าพิกัดจีพีเอสอยู่ในอีสานหรือไม่