USING MACHINE LEARNING FOR UNSUPERVISED MARITIME ROUTE DISCOVERY
SYNCHROMODAL IT – REAL TIME AND BIG DATA
ANDREJ DOBRKOVIC
USING MACHINE LEARNING FOR UNSUPERVISED MARITIME ROUTE DISCOVERY

AGENDA

- Data streams
- Synchronomodal Logistics
  - Unscheduled arrival problem
- Unsupervised maritime route discovery
- Algorithms
  - DBSCAN
  - Modified ant colony optimization
  - Genetic algorithm
  - Hybrid approach
DATA STREAMS
CONCEPT

- Data arrives at sequence of items
- At high speed
- Forever
- Cannot store it all
- Cannot go back or slow down
- Evolving
DATA STREAMS
AXIOMS

- Only one, $t$-th item available at time $t$ only
- Each item can only be accessed once
- Limited resources (memory / time)
- Incorporate forgetting mechanisms
- Able to provide answers at any time
- The stream evolves over time
DATA STREAMS
DATA NEVER SLEEPS
DATA STREAMS
DATA NEVER SLEEPS
SYNCHROMODAL LOGISTICS
PROJECT MOTIVATION

- The need for unified European logistics network
- Increase the efficiency and sustainability of logistics services
- Motivate shippers and 4PLs to switch to **synchromodal way of working**
  - Decisions about the next leg of the route are made as late as possible
  - Optimal flexibility in routing
  - Improved quality of service and sustainability
SYNCHROMODAL LOGISTICS
SYNCHROMODAL IT PLATFORM
SYNCHROMODAL LOGISTICS
CONSOLIDATION FOR LAST MILE REDUCTION
SYNCHROMODAL LOGISTICS
CONSOLIDATION FOR LAST MILE REDUCTION

\[
\begin{align*}
\min Z \\
\text{s.t.} \\
Z &= \sum_{t \in T} \sum_{D' \subseteq D} \left( C_{D'} \cdot \prod_{d' \in D'} y_{d',t} \cdot \prod_{d'' \in D_t \setminus D'} (1 - y_{d'',t}) \right) + \sum_{t \in T} \sum_{f \in F} \sum_{d \in D} (B_d \cdot L_{f,d} \cdot v_{f,t}) \\
\sum_{f \in F} &\geq Q, \quad \forall t \in T \\
\sum_{t \in T} &\leq R_f, \quad \forall f \in F \\
(L_{f,d} \cdot x_{f,t}) - M_d \cdot y_{d,t} &\leq 0, \quad \forall d \in D_t, t \in T \\
D_t &= \{d : L_{f,d} = 1 \text{ and } R_f \leq t, \forall d \in D, f \in F\}, \quad \forall t \in T \\
x_{f,t} &\in \{0,1\}, \quad \forall f \in F, t \in T \\
y_{d,t} &\in \{0,1\}, \quad \forall d \in D_t, t \in T \\
v_{f,t} &\in \{0,1\}, \quad \forall f \in F, t \in T \\
Z &= \sum_{t \in T} \sum_{D' \subseteq D} (C_{D'} \cdot w_{D',t}) + \sum_{t \in T} \sum_{f \in F} (B \cdot S_f \cdot v_{f,t}) \\
w_{D',t} - y_{d',t} &\leq 0, \quad \forall D' \subseteq D_t, d' \in D_t, t \in T \\
w_{D',t} + y_{d',t} &\leq 1, \quad \forall D' \subseteq D_t, d' \in D_t \setminus D', t \in T \\
w_{D',t} + (|D'| - 1) - \sum_{d' \in D'} y_{d',t} + \sum_{d'' \in D \setminus D_t} y_{d'',t} &\geq 0, \quad \forall D' \subseteq D_t, t \in T
\end{align*}
\]
UNCERTAINTIES
UNSCHEDULED ARRIVAL OF DEEP SEA VESSEL

- Inaccurate planning
- Weather conditions
- Tides
- Skipper behavior

- Deep sea vessel has priority over barge at terminal
- Alternative: send trucks
HOW DO PLANNERS COPE WITH PROBLEM?
USE AIS DATA TO ESTIMATE ARRIVALS

- Automatic Identification System
  - Used for collision avoidance
  - Ship data can be received by any AIS radio receiver
- Manually check AIS based web pages and … guess!

ARRIVAL PREDICTION
STATE OF THE ART

- Considerable research done in estimating trajectory
  - Short-term prediction only (less than 1 hour)
  - Collision avoidance / anomalous behavior tracking
  - Unreliable results with increasing time horizon
- Short term prediction – vessel motion patterns
- Long term prediction – route extraction

<table>
<thead>
<tr>
<th>Objective</th>
<th>Timeframe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>short (less than 1 hour)</td>
</tr>
<tr>
<td></td>
<td>long (more than 1h)</td>
</tr>
<tr>
<td>A vessel motion patterns</td>
<td>(Wijaya &amp; Nakamura, 2013)</td>
</tr>
<tr>
<td></td>
<td>(Wang et al., 2013)</td>
</tr>
<tr>
<td></td>
<td>(Hornauer &amp; Hahn, 2013)</td>
</tr>
<tr>
<td></td>
<td>(Laxhammar et al., 2009)</td>
</tr>
<tr>
<td></td>
<td>(Ristic et al., 2008)</td>
</tr>
<tr>
<td>B route extraction</td>
<td>(Vespe et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>(Talavera et al., 2013)</td>
</tr>
<tr>
<td></td>
<td>(Lei et al., 2011)</td>
</tr>
<tr>
<td>A + B route extraction and motion patterns</td>
<td>(Liu &amp; Chen, 2014)</td>
</tr>
<tr>
<td></td>
<td>(Pallotta et al., 2013)</td>
</tr>
</tbody>
</table>
- Use unsupervised learning to extract waypoints – nodes
- Connect nodes with probability of visiting another – edges
- Use constructed graph to estimate vessel’s destination

Use data as available to LSPs:
- The North sea
- Average 693 raw AIS messages each second
- Good coastal AIS coverage
- Limited AIS reception from open sea
- Store data in temporary buffer
- Process buffer every 5 – 7 minutes
ALGORITHMS

- DBSCAN
- Modified ant colony optimization
- Genetic algorithm
- Hybrid approach
Discover clusters of arbitrary shape
Handle noise
One scan
Dependent on two parameters:
  - Distance ($\varepsilon$)
  - Number of points (minPoints)
DBSCAN
RESULTS / CONCLUSIONS

- Fast
- Can handle noise
- Varying extraction quality
  - Density issues
- Cannot handle streaming data
DBSCAN PROBLEMS

- Varying densities
- High dimensional data

(MinPts=4, Eps=9.75)

(MinPts=4, Eps=9.92)
MODIFIED ANT COLONY OPTIMIZATION

CHARACTERISTICS

- Metaheuristic optimization algorithm (swarm intelligence)
- Search optimal path in a graph based on ant behavior
  - “Pheromone trail”
  - Pheromone fade over time
  - Discover routes and waypoints based on pheromone density
MODIFIED ANT COLONY OPTIMIZATION
ADAPTING TO AIS DATA STREAM
MODIFIED ANT COLONY OPTIMIZATION
RESULTS
MODIFIED ANT COLONY OPTIMIZATION

CONCLUSIONS

- Fast
- Handles large data streams well
- Can provide answer anytime
- Forgetting mechanism
- Sensitive to parameters; problem with different density areas
- Can misinterpret noise for route
GOOD FOR SOLVING SEMI-STRUCTURED AND UNSTRUCTURED PROBLEMS

- Quick convergence towards optimal solution
- Produce “good solution”, but not optimal solution
GENETIC ALGORITHM
ADAPTING TO AIS DATA STREAM
Experiment results / conclusions:
- Efficient WP discovery on high density routes
- Handles noise and streaming data well
- Evolve with data
- Prioritizes high density routes and ignores low density ones
- Slow
**HYBRID APPROACH**

**CHARACTERISTICS**

- Combines the strongest features of the previous algorithms

1. Subdivide the problem space using QuadTree into sub regions
2. Assess potential of each sub region to contain waypoints
3. Apply DBSCAN on good candidates
4. Use “pheromone trail” on discovered waypoints
HYBRID APPROACH
ADAPTING TO AIS DATA STREAM
HYBRID APPROACH
RESULT
HYBRID APPROACH

CONCLUSIONS

- Density tolerant (extracts WP regardless of route density)
- Fast
- Noise tolerant
- Somewhat intolerant to “dark areas” (areas from which periodically no AIS data is received)
EVALUATION CRITERIA

- Extraction quality
- Algorithm efficiency
- Traffic density tolerance
- Noise tolerance
- Blank region tolerance
EXTRACTION QUALITY
TEST CASE

UNIVERSITY OF TWENTE.
EXTRACTION QUALITY
RESULTS

DBSCAN 1    DBSCAN 2    DBSCAN 3

Gen. alg.    Mod. ACO    Hybrid approach
# Extraction Quality Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Route</th>
<th>Intersections</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBSCAN 1</td>
<td>25%</td>
<td>50%</td>
</tr>
<tr>
<td>DBSCAN 2</td>
<td>25%</td>
<td>50%</td>
</tr>
<tr>
<td>DBSCAN 3</td>
<td>19%</td>
<td>25%</td>
</tr>
<tr>
<td>Genetic Algorithm</td>
<td>69%</td>
<td>75%</td>
</tr>
<tr>
<td>Modified ACO</td>
<td>0%</td>
<td>50%</td>
</tr>
<tr>
<td>Hybrid Approach</td>
<td>90%</td>
<td>75%</td>
</tr>
</tbody>
</table>

![Bar chart showing extraction quality results for different methods.](chart.png)
## ALGORITHM EFFICIENCY

### RESULTS

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>DBSCAN</th>
<th>Genetic Algorithm</th>
<th>Mod. ACO</th>
<th>Hybrid Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Execution time (sec.)</td>
<td>10 *</td>
<td>3600 **</td>
<td>10 *</td>
<td>20 - 43</td>
</tr>
</tbody>
</table>
### EVALUATION AND CONCLUSIONS

<table>
<thead>
<tr>
<th>Criteria</th>
<th>DBSCAN</th>
<th>GA</th>
<th>Modified ACO</th>
<th>Hybrid Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraction quality</td>
<td>+ / -</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Algorithm efficiency</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Traffic density</td>
<td>-</td>
<td>+ / -</td>
<td>+</td>
<td>+ / -</td>
</tr>
<tr>
<td>Noise tolerance</td>
<td>+</td>
<td>+</td>
<td>+ / -</td>
<td>+</td>
</tr>
<tr>
<td>Blank region tolerance</td>
<td>-</td>
<td>+ / -</td>
<td>+ / -</td>
<td>+ / -</td>
</tr>
</tbody>
</table>
FUTURE WORK

- Validation
  - Enhance evaluation metrics (graph similarity using random walks)
  - Increase test scenarios
  - Extend testing interval (months instead of days)

- Algorithm improvement
  - DBCAN – recursive depth search
  - Mod. ACO – image recognition
  - Gen. alg. – incremental evolution
THANK YOU!
ANDREJ DOBRKOVIC
IEBIS – UNIVERSITY OF TWENTE
a.dobrkovic@utwente.nl