Algorithmic Decision Theory and the Port Reopening Scheduling Problem

Fred S. Roberts
Director of CCICADA
Rutgers University
Vulnerability of Ports

• Considerable effort to introduce data-science-based methods to make our ports safer.
• Other CCICADA efforts on container inspection are part of that.
• Those efforts concentrate on inspection of cargo.
• The project I will describe is concerned with ways in which ports might be shut down in part or entirely and then reopened.
Vulnerability of Ports

• Ports might be shut down by terrorist attacks, natural disasters like hurricanes or ice storms, strikes or other domestic disputes, etc.

• Project themes:
  – How do we design port operations to minimize vulnerability to shut down?
  – How do we reschedule port operations in case of a shutdown?
Algorithmic Decision Theory

• This project is related to a larger thrust on “algorithmic decision theory”

• Today’s decision makers in fields ranging from engineering to medicine to homeland security have available to them:
  – Remarkable new technologies
  – Huge amounts of information
  – Ability to share information at unprecedented speeds and quantities
Algorithmic Decision Theory

• These tools and resources will enable better decisions if we can surmount concomitant challenges:
  – The massive amounts of data available are often incomplete or unreliable or distributed and there is great uncertainty in them
Algorithmic Decision Theory

• These tools and resources will enable better decisions if we can surmount concomitant challenges:
  – Interoperating/distributed decision makers and decision-making devices need to be coordinated
  – Many sources of data need to be fused into a good decision, often in a remarkably short time
Algorithmic Decision Theory

• These tools and resources will enable better decisions if we can surmount concomitant challenges:
  – Decisions must be made in dynamic environments based on partial information
  – There is heightened risk due to extreme consequences of poor decisions
  – Decision makers must understand complex, multi-disciplinary problems
Algorithmic Decision Theory

• In the face of these new opportunities and challenges, ADT aims to exploit algorithmic methods to improve the performance of decision makers (human or automated).
• Long tradition of algorithmic methods in logistics and planning dating at least to World War II.
• But: algorithms to speed up and improve real-time decision making are much less common.
Reopening a Port After Shutdown

- Shutting down ports is not unusual – e.g., hurricanes
- Scheduling and prioritizing in reopening the port is often done very informally
- Improving on existing decision support tools for port reopening could allow us to take many more considerations into effect
- Can modern algorithmic methods of ADT help here?
Manifest Data

• Part of the solution to the port reopening problem: Detailed information about incoming cargo:
  – What is it?
  – What is its final destination?
  – What is the economic impact of delayed delivery?

• A key is to use container *manifest data* to estimate economic impact of various disaster scenarios & understand our port reopening requirements
Data Description

- We obtained from CBP one month’s data of all cargo shipments to all US ports
- Jan 30, 2009 – Feb 28, 2009

Description
- Foreign port (origin)
- Domestic port (destination)
  - Aggregation
  - Item description
  - Item count
  - Inconsistencies
Data Description

• Data has errors and ambiguities
• Does 150 waters mean 150 bottles of water or 150 cases of bottles of water?
• What does “household goods” mean?

• Still, there are things we can do with the data.
Mining of Manifest Data

• Separate effort: Predict risk score for each container
  – Quantify the likelihood of need for inspection
  – Based on covariates/characteristics of a container’s manifest data.

• Methods:
  – We are developing machine learning algorithms to detect anomalies in manifest data.
  – Text mining on verbiage fields.
  – Logistic regression with LASSO.
    • Simulation study conducted suggests that the LASSO regression approach is an effective tool for processing information in the manifest data
Visualization Tools Applied to Manifest Data

• Visualizing data can give us insight into interconnections, patterns, and what is “normal” or “abnormal.”

• Visualization is part of another effort, but similar methods can help with the port reopening problem.

• Our visual analysis methods are based on tools originally developed at AT&T for detection of anomalies in telephone calling patterns – e.g., quick detection that someone has stolen your AT&T calling card.

• The visualizations are interactive so you can “zoom” in on areas of interest, get different ways to present the data, etc.

• Work of James Abello and Tsvetan Asamov
Visualization Tools for Risk Assessment/Anomaly Detection

- For port $p$, a vector $contents[p]$ gives the number of items of each kind of commodity shipped out of port $p$ in a given time period.
- We devise *similarity measures* between ports $p$ and $q$ as a function of the dot product of their contents vectors.
- $Contents[p,q]$ gives the number of items of each kind of commodity shipped from port $p$ to port $q$ in a given time period.
- We represent such vectors using edge-weighted, labeled graphs that can be visualized using our software.
General View of Port to Port Traffic

Color-coded connections represent number of shipments
Shanghai, LA, Newark, Singapore

Vertex Size encodes number of shipments
Zooming into Shanghai (gray)

Zooming into a vertex gives more data about traffic
Contents To Port Pairs

Vertices are KeyWords and Port Pairs (color coded by degree), Edges encode number of containers (or shipments) with that keyword for the corresponding Port Pair.
Contents To Port Pairs (cont)
(Vertices color coded by WeightRatio)
Temporal Evolution of Manifest Data

Fix a commodity.
Each vertex represents all shipments from foreign to US ports on a given day.
Cluster by similarity. Notice how all Tuesdays and Wednesdays are well clustered.
Can also Cluster by Ports

Note similarity, e.g., Cincinnati, OH and Brunswick, GA
Resilience Modeling

• If a port is damaged or closed, immediate problem of rerouting some or all incoming vessel traffic – if the reopening will be delayed for awhile.

• Also: problem of prioritizing the reopening of the port – and deciding whether and how to reorder ships’ arrivals/unloading

• These problems can be subtle.
  – Ice storm shuts down port
  – Maybe priority is unload salt to de-ice. It wasn’t a priority before.
Resilience Modeling: One Port

• *Problem:* Reschedule unloading of queued vessels.
Resilience Modeling: One Port

- **Problem: Reschedule unloading of queued vessels.**
  - Done by consult with *shippers* and their priorities
  - Also consult with key *government agencies* to target priority goods or shipments
  - Take into account potential *spoilage* of cargo
  - Take into account acute *shortage* of key items: food, fuel, medicine, etc.
  - Thus: *Many variables* to take into account and juggle
  - Want *systematic methods*; don’t want one stakeholder to feel that only other stakeholders’ views were taken into account.
  - Methods of algorithmic decision theory can help
  - So far, just beginning to define the problem and identify the key challenges in developing decision support tools
Resilience Modeling: One Port

- **Problem:** Reschedule unloading of queued vessels.
- Think of a ship as corresponding to a vector \( x = (x_1, x_2, \ldots, x_n) \) where \( x_i \) is quantity of \( i^{th} \) good.
- Assume all \( x_i \) are integers.
- Suppose for simplicity that each ship takes the same amount of time to unload. Then we can give each ship a *timeslot* for unloading.
- The port’s capacity determines how many ships can be scheduled at a given timeslot.
- Suppose we require \( d_i \) units of good \( i \) by timeslot \( t_i \).
- \( d = (d_1, d_2, \ldots, d_n) \), \( t = (t_1, t_2, \ldots, t_n) \).
- In practice, require \( d_{i1} \) units of good \( i \) by timeslot 1, \( d_{i2} \) units of good \( i \) by timeslot 2, etc. Disregard this.
- **How do we assign ships to timeslots?**
Resilience Modeling: One Port

There are some *subtleties*:

- The manifest data is unclear. If \( i \) is water, \( x_i = 150 \) could mean 150 bottles of water or 150 cases of bottles of water.
- The manifest data is unclear: Descriptions like “household goods” are too vague to be helpful.
- Different goods have different *priorities*. For example, not having enough food, fuel or medicine is much more critical than not having enough bottles of water.
- Let \( p_i = \) the priority assigned to good \( i \), with \( p = (p_1, p_2, \ldots, p_n) \).
Resilience Modeling: One Port

There are some *subtleties*:

- There are *penalties for late arrivals of goods*.
- Sometimes there are *even penalties for early arrivals* (storage space issues)
- The penalty can depend on the priorities.
Resilience Modeling: One Port

We encountered a similar problem in working for the Air Mobility Command of the US Air Force.

- Fly soldiers from point A to point B
- Each has desired arrival time
- Getting a general there late is worse than getting a private there late

Similar problems also arise in machine scheduling.
We speak of machine scheduling with earliness and tardiness penalties = “just in time scheduling”
Resilience Modeling: One Port

• We have been looking at the simplified problem with all $d_i = 1$ (we demand exactly one unit of each good).
• Then we can talk about the first time a ship carrying good i is scheduled, $S_i$.
• Let $S = (S_1, S_2, \ldots, S_n)$
Resilience Modeling: One Port

- We are looking at a number of different objective functions \( F(S,t,p) \) that need to be optimized.
- Let **tardiness** \( T_i = \max\{0, S_i - t_i\} \), **earliness** \( E_i = \max\{0, t_i - S_i\} \).
- For example:
  - \( F(S,t,p) = \sum p_i T_i + \sum p_i E_i \)
  - \( F(s,t,p) = \sum p_i T_i \)
  - \( F(s,t,p) = \sum h(p_i) T_i \)
  - \( F(s,t,p) = \max\{h(p_1) T_1, h(p_2) T_2, \ldots, h(p_n) T_n\} \)
Resilience Modeling: One Port

- A very special case:
  - Only one ship can be unloaded at a time
  - Each ship carries only one kind of good
  - All goods have the same desired arrival time, ♦

- Let \( F(S,t,p) = \bigvee h(p_i)T_i \), \( h(p_i) \) increasing in \( p_i \).

- Then a **greedy algorithm** gives the optimal unloading schedule: Schedule the highest priority good for first arrival, then the second highest priority good,
Resilience Modeling: One Port

• Now allow more than one good per ship, but still:
  – Only one ship can be unloaded at a time
  – All goods have the same desired arrival time, ♦
• Let $F(S,t,p) = \sum h(p_i)T_i$, $h(p_i)$ increasing in $p_i$.
• Ship 1 has $x = (1,0,0,0)$, ship 2 has $x = (0,1,0,0)$, ship 3 has $x = (0,0,1,1)$
• ♦ = 2
• $p_1 > p_2 > p_3 > p_4$
• Now greedy algorithm would take ship 1 first, then ship 2, then ship 3.
• The penalty for this schedule is $h(p_3) + h(p_4)$.
• But: scheduling ship 1 first, then ship 3, then ship 2 has penalty $h(p_2)$, which might be smaller.
• The problem is subtle and even in this special case not simple to solve.
Resilience Modeling: One Port

• Now allow exactly two goods per ship
  – The port has capacity for only one ship at a time.
• This translates into a graph theory problem where ships correspond to edges, goods to vertices, and we want to order the edges in such a way that every vertex gets assigned a timeslot that corresponds to the earliest timeslot of any edge it belongs to.
Resilience Modeling: Simplifications

- All desired amounts are one unit, i.e., $d_i = 1$, all $i$.
- Reopened port has limited capacity of one ship per timeslot.
- All goods have the same desired arrival time ♦
- All goods have only one desired arrival time rather than portion desired by time 1, some by time 2, etc.
- All ships have same unloading time.
- All ships are ready to dock without delay.
- There is no problem storing unloaded but not urgently-demanded goods.
- Each ship has only one kind of good.

Even making all or most of these assumptions leads to a complex scheduling problem.
Resilience Modeling: Rerouting to Nearby Ports

• Problem: If a port can’t be reopened soon, incoming ships must go to other ports.
• As a general rule, where they go is left to shippers/vessel operators
• However, can we develop a decision support tool that will allow us to provide guidance to shippers and take into account the need to deliver critical supplies?
• Goals: minimize economic impact of delay and security impact of delay in delivery of critical supplies.
• Start with one nearby port; then try two ports
• This work is planned.
• Step one is to identify priorities for where goods are to be delivered.
Determining the Priorities

• How do we determine the priorities as to where different goods are to be delivered?
• One approach: each stakeholder (government, port operators, shippers) provides their priorities and some *consensus or voting procedure* is used to “average” them.
Determining Desired Times for Unloading

- Explore “bidding system” for setting times to unload vessels.
- After government or central entity sets desired times for delivery of critical products, companies receiving shipments make bids for earliest arrival dates.
- Problem is complicated by mixed collection of goods in any container.
- Based on priorities and bids, find ways to do “optimal” rerouting.
Determining Desired Times for Unloading

- The “bidding” system gets into the mathematical analysis of auctions.
- A topic of a great deal of research in mathematics and computer science.
- Information technology allows complex auctions with a huge number of bidders.
- *Bidding protocols maximizing expected profit can be extremely difficult to compute.*
Determining Desired Times for Unloading

• Multiple goods are to be auctioned off.
• In practice, you submit bids for combinations of goods.
• This leads to NP-complete allocation problems.
• Might not even be able to feasibly express all possible preferences for all subsets of goods.
• Then: Determining the winner of an auction can be extremely hard. (Rothkopf, Pekec, Harstad)
Determining Desired Times for Unloading

• In these complicated “combinatorial auctions,” we need to elicit preferences from all “players” for all plausible combinations of items in the auction.

• Similar problem arises in optimal bundling of goods and services.

• Elicitation requires exponentially many queries in general.

• Thus, *bidding procedures to aid in reopening closed ports lead to challenges for modern methods of ADT.*
Collaborators

• **Data Visualization**
  - James Abello, Tsvetan Asamov

• **Decision Support for Reopening a Port**
  - Paul Kantor, Endre Boros, Tsvetan Asamov, Emre Yamangil

• **Bidding Systems**
  - Paul Kantor, Aleksandar Pekec

Thank You!!