Sheaf Theory: The Mathematics of Data Fusion

Michael Robinson

SIMPLEX Program

© 2015 Michael Robinson
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
Acknowledgements

Reza Ghanadan (DARPA/DSO)

Collaborators:

• Cliff Joslyn & Emilie Hogan
  – Semantic technologies
  – Discrete mathematics

• Chris Capraro
  – Signal processing system design
  – High performance computing

• Other SIMPLEX teams: Georgetown University, Baylor Medical College, Vanderbilt University, MIT, ...
Logistics

- What's where?
- Discussion areas
- Wireless access
- Streaming, archiving
  - Lecture “ground rules”
- How the interactive sessions work
# Tutorial schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Tuesday, August 25</th>
<th>Wednesday, August 26</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00-9:00am</td>
<td>Registration</td>
<td></td>
</tr>
<tr>
<td>9:00-9:15am</td>
<td>Opening remarks</td>
<td></td>
</tr>
<tr>
<td>9:15-10:00am</td>
<td>Sheaf theory: the Mathematics of Data Fusion</td>
<td>Categorification and Chain Complexes</td>
</tr>
<tr>
<td>10:15-11:00am</td>
<td>Topological Features</td>
<td>Constructing and Categorifying</td>
</tr>
<tr>
<td>11:15-12:00pm</td>
<td>What is Topology?</td>
<td>Computing Topological Features</td>
</tr>
<tr>
<td>12:00-1:30pm</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>1:30-2:15pm</td>
<td>What is a Sheaf?</td>
<td>Sheaf Cohomology and its Interpretation</td>
</tr>
<tr>
<td>2:30-3:15pm</td>
<td>Building Complexes and Sheaves</td>
<td>Computing Homology and Cohomology</td>
</tr>
<tr>
<td>3:30-4:15pm</td>
<td>Data Structures as Sheaves</td>
<td>How do we Deal with Noisy Data?</td>
</tr>
</tbody>
</table>
Tutorial objectives

- What are sheaves?
  - The “local to global” viewpoint
- Encoding existing data into sheaves
  - “Sheafification”
- Data analytic capabilities enabled by sheaves
  - “Sections,” “cohomology”
- Practice analyzing sheaves in software
  - “Persistence,” “local homology”
- Interpret this analysis into the context of the data
The SIMPLEX project

**Problem**: Heterogeneous data fusion has been stymied by the lack of effective quantitative – qualitative fusion techniques

**Key points**:  

- *Sheaf-based* heterogeneous fusion is the sole canonical mathematical framework for multi-modal, multi-INT integration systems  
- Demonstrate the framework by showing cross-modality inference using *topological invariants*
Heterogeneous integration systems

A mathematical framework for heterogeneous integration should

- Be general enough to accurately represent the data in all its richness
- Be capable of summarization of the data into important, actionable features

**Theorem:** These two properties lead directly to a sheaf-based data representation

- Any other representation recapitulates sheaf theory
Benefits of sheafification

Solid theoretical grounding yields powerful inference capabilities:

• **Globalization** of sensor data
  – Sheaves are the correct theory of data fusion
  – Track generation and disambiguation

• **Topological sensor retasking**
  – Detection of misinformation and deception
  – Detection of “loopy” belief propagation
  – Removal of inferential vulnerabilities and self-inconsistencies
Steps of heterogeneous fusion

1. Form the *base topological space*: multiway interactions between data sources

2. *Sheafify*: build the model of relations between data sources

3. *Categorify*: place the data streams in vector spaces to aid in computation and analysis

4. Compute *cohomology*: globalize the data to find robust invariants
Mathematical dependency tree

- Linear algebra
  - Manifolds
  - Calculus
  - Homology
  - de Rham cohomology (Stokes' theorem)
- Sheaves
  - Sheaf cohomology
  - Cellular sheaves
  - CW complexes
  - Abstract Simplicial Complexes
  - Simplicial Complexes
- Topology
  - Set theory
  - Lecture 2
- Lectures 3, 4
- Lectures 5, 6
- Lectures 7, 8
Session objectives

- Provide an overview of topological data fusion

- When does a problem involve “topological” features?

- When do sheaves add a new capability beyond topology?
What are topological features?

- Clusters
- Holes
- Voids
- Boundaries
- Dimension
- Local “model” spaces
Sheafification

Getting your data into a sheaf

(and what's that anyway?)
What is a sensor?

- Anything that provides a measurement
- It could be something that records an image...
What is a sensor?

- Anything that provides a measurement
- … or a pixel …
What is a sensor?

• Anything that provides a measurement
• … or even some text

“21 train cars, two of which are yellow”
Sheaf: a heterogeneous sensor system

A sensor transforms a physical region into data

“Physical” sensor footprints | Sensor data space

Optical | SIGINT

Images | Ellipses
Sheaf: a heterogeneous sensor system

Optical → SIGINT → Summarization → “Physical” sensor footprints

Images → Ellipses

Sensor data space
Sheaf: a heterogeneous sensor system

This construction – the data together with the transformations – is a sheaf

Optical

SIGINT

“Physical” sensor footprints

Sensor data space

Fused targets

Images

Ellipses

Detections

Globalization

Michael Robinson
Globalizing data in a sheaf
Multi-INT tracking

- Combine measurements of a scene from **multiple sensors** returning **arbitrary data types** to make inferences about the scene.

HUMINT observation

SIGINT observation

AMTI track
Topological fusion

- Topologize the space of sensors, abstractly
Topological fusion

- **Encode** the data collected by each sensor
Topological fusion

- Express **consistency requirements** for adjacent sensors
Globalization of data

- Finding **maximally consistent information** through this network amounts to **tracking**
Multi-INT tracking: input data

Position

SIGINT

HUMINT

Image

SIGINT

Image

HUMINT
Multi-INT tracking: models

Regions describing where targets might be found after detection

Minor technical point: this is actually a cosheaf
Multi-INT tracking: tracks

Position

SIGINT

HUMINT

Image

SIGINT

HUMINT

Time

Michael Robinson
Single-INT demonstration

- Processing: pulse-wise CFAR followed by centroid
- Moving receiver, stationary transmitter
- 44.1 kHz sample rate
- 1500 range samples per pulse

Collection system:
Netbook
Wired microphone
10W Laptop speakers

Fine structure of near-in clutter

Crossings in range

Drop-outs due to shadowing

Strong “tracks” of specular multipath from walls

PRI

Sample

0 50 100 150 200 250

Pulse

0 500 1000 1500
Single-INT tracker results

Kalman filter

Substantially reduced confounding

Cosheaf tracker
Detecting defects in the data
Misleading fused data: Loops

- Sheaves are sensitive to *topological features*
  - These are features that are robust to *large* perturbations
  - Clusters, loops, voids, etc.

- The presence of loops in the sensor structure indicates
  - Coverage holes
  - Possibility of spurious information going unnoticed
“Weather Loop” a simple model

<table>
<thead>
<tr>
<th>Sensors/Questions</th>
<th>Rain? (R)</th>
<th>Humidity % (H)</th>
<th>Clouds? (L)</th>
<th>Sun? (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>News (N)</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Weather Website (W)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Rooftop Camera (C)</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Twitter (T)</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Question: Can misleading globalized information be detected?
Misleading globalized data

News: “There are some sun showers in the area”
Weather website: 100% chance of rain; Humidity is at 75%
Rooftop camera: Only blue sky in view
Twitter: “@weatherguy Dude it’s so muggy out, but at least I see some blue sky!”

- Each edge is consistent – so this is globalized
- No global consistency check (no higher dimensional face), so there can be conflicting information
  - For instance: rain without clouds

Under the right conditions, this can be detected algorithmically
Next up...

- Interactive session: Topological Features
- Next lecture: What is Topology?