MACHINE LEARNING FROM MULTIMODAL DATA

Andrea Burns
Dr. Waheed Bajwa
Dr. Mark Pierce
RECAP: MOTIVATION

- Investigating the impact of using multimodality in the process of image classification
- Our form of multimodality: hyperspectral imaging
  - Each spectral band forms a different mode
AVIRIS Indian Pines Data Set: Hyperspectral imaging

<table>
<thead>
<tr>
<th>Class #</th>
<th>Class</th>
<th># Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alfalfa</td>
<td>46</td>
</tr>
<tr>
<td>2</td>
<td>Corn-notill</td>
<td>1428</td>
</tr>
<tr>
<td>3</td>
<td>Corn-mintill</td>
<td>830</td>
</tr>
<tr>
<td>4</td>
<td>Corn</td>
<td>237</td>
</tr>
<tr>
<td>5</td>
<td>Grass-pasture</td>
<td>483</td>
</tr>
<tr>
<td>6</td>
<td>Grass-trees</td>
<td>730</td>
</tr>
<tr>
<td>7</td>
<td>Grass-pasture-mowed</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>Hay-windrowed</td>
<td>478</td>
</tr>
<tr>
<td>9</td>
<td>Oats</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>Soybean-notill</td>
<td>972</td>
</tr>
<tr>
<td>11</td>
<td>Soybean-mintill</td>
<td>2455</td>
</tr>
<tr>
<td>12</td>
<td>Soybean-clean</td>
<td>593</td>
</tr>
<tr>
<td>13</td>
<td>Wheat</td>
<td>205</td>
</tr>
<tr>
<td>14</td>
<td>Woods</td>
<td>1265</td>
</tr>
<tr>
<td>15</td>
<td>Buildings-Grass-Trees-Dives</td>
<td>386</td>
</tr>
<tr>
<td>16</td>
<td>Stone-Steel-Towers</td>
<td>93</td>
</tr>
</tbody>
</table>
Data set includes:

- Image of Indian Pines is 145 x 145
- Data is 145 x 145 x 200, 200 spectral bands
- Each band is 10µm

Classified the 12 largest classes (dropping the 4 smallest to improve accuracy)

- SVM with one against rest multiclass-classification
- 80% training, 20% testing – randomly shuffled and split
RESULTS

- Calculated accuracy of classification for 20-, 40-, 50-, 100-, 150-, and 200-band data, as well as with different ranges for each of these # of bands.
ACCURACY RESULTS

- **20 bands** ~35-54%
  - Best range: 0.2-0.4µm, 54.975%

- **40 bands** ~49-60%
  - Best range: 1.45-1.99µm, 60.338%

- **50 bands** ~54-62%
  - Best range: 1.2-1.89µm, 62.761%

- **100 bands** ~68-71%
  - Best range: 0.2-1.2µm, 71.341%

- **150 bands** ~74-79%
  - Best range: 0.2-1.89µm, 79.72%

- **200 bands** ~79-83% (0.2-2.39µm)
The number of spectral bands affects accuracy
- The more bands used – the higher accuracy

The range of spectral bands affects accuracy
- Certain ranges (of the same number of bands) perform slightly better than others
- Suggests that different bands can extract more useful information from an image for classification than others
The hay-windrowed class consistently performed best. It was shown that even as the number of bands decreased, it had high accuracy.

Additionally, the ranges excluding 1.89-2.39µm typically performed better amongst the same # of bands. Indicates some ranges can identify unique qualities of each class more so than others.
Classification of liquids: 12 classes in total

Empty, water, hydrogen peroxide, rubbing alcohol, milk, soy milk, lemonade vitamin water, cranberry juice, orange juice, Ginger Ale, Coca Cola, seltzer water

Visible light and UV light camera examples
Final data will be 640 x 480 x 12 cuboid per class

Resolution and pixel size is not consistent between cameras
- Need to match to lowest resolution
- Covert pixel size to lowest scale (8 bit int)
- Account for alignment issues

12 Bands: Achieved with several cameras and filters
- UV light
- Visible light
- NIR/SWIR (4 bands w/filters)
- MWIR (3 bands w/filters)
- LWIR
How will overlap and discretization of the spectral ranges affect my results?

Will the results be consistent with my analysis of the AVIRIS Indian Pines data set?

Can multimodality help distinguish liquids of the same or similar color?
FINAL GOALS

- Process data (i.e. same resolution, pixel value range, and alignment)

- Classify data and analyze the results, potentially answering these research questions

- If time allows, perform the same process again with another hyperspectral data set (potentially tensor data of different dimensions)
I would like to acknowledge and thank the National Science Foundation for funding this research. This work has been supported by NSF grant CCF-1559855.

An additional thanks to my mentor Dr. Waheed Bajwa, Dr. Mark Pierce, Lazaros Gallos, and the kind, informative graduate students of the INSPIRE lab for helping me succeed.