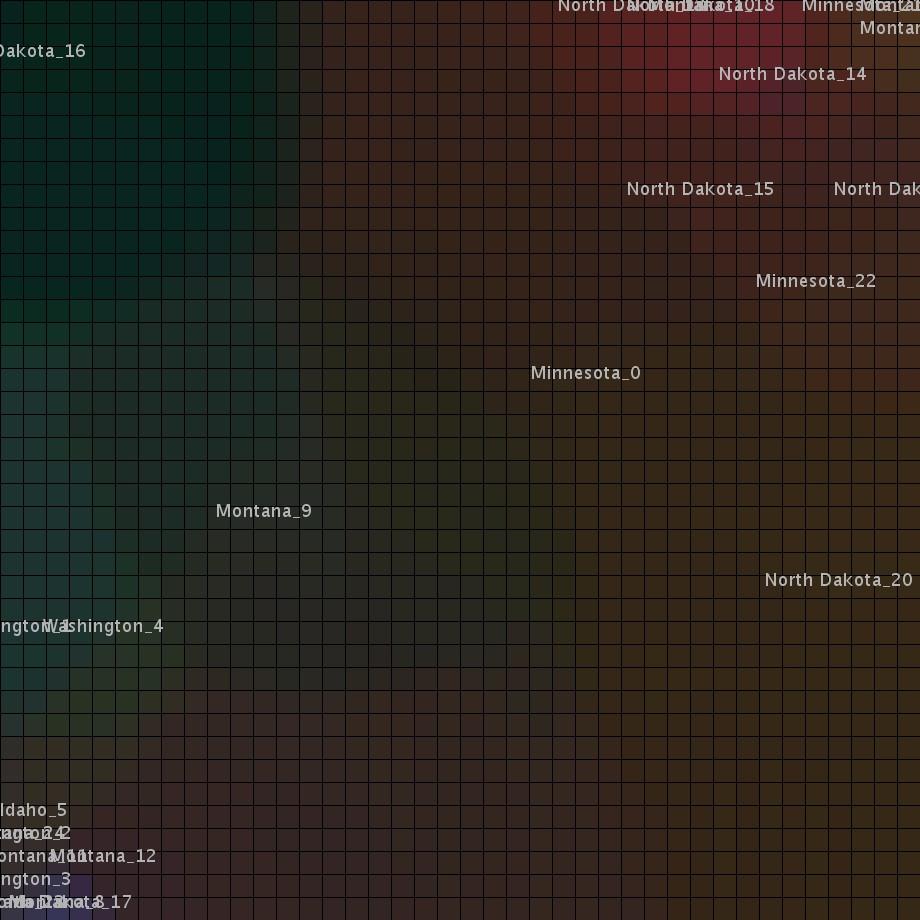
Ryan McGrath

UMBC 636 Data Visualization

Visualization Implementation



*A self organizing map visualization of the rate of breast cancer, percentage of residents in poverty, and percentage of residents unemployed in various locations. The rate of breast cancer is mapped to the red color channel, unemployment is mapped to the green color channel, and percent poor is mapped to the blue color channel. Locations exhibit clustering near locations that have related rates for each field and tend to move towards colored regions which strongly identify the dominant field. One observation is that North Dakota\_16 in the upper left region has a high unemployment rate, but low breast cancer and low poverty rate. Locations clustered in the upper right corner tend to have a high breast cancer rate.*

The Self Organizing Map (SOM) implementation accepts a comma separated values (CSV) file as input and maps values to a set of sample vectors. The user controls characteristics of the visualization implementation by changing global variables in the configurable options section at the start of the code. The height, width, grid\_height, and grid\_width variables control the size and display of the visualization. Iterations determines how many steps the algorithm takes, and the learning constant affects the magnitude of the decreasing learning rate. The sample count determines how many rows should be read from the CSV file. The user can select three fields from each sample using field1, field2, and field3. Three values are used since this implementation maps each field to a RGB color channel.

The SOM algorithm breaks down roughly into the following steps:

Select a sample

Get best matching node to the sample

Scale neighbors to be like the sample to a decreasing extent within a decreasing radius

Iterate

Different descriptions of SOM recommend different algorithms and techniques to implement these steps. To ease implementation, most steps were constructed in the most straightforward manner. Samples are chosen in order, the neurons are initialized to a random value for each RGB channel, the best matching unit is the least value of a three dimensional euclidian distance, and the radius decreases using a linear function found through trial and error. For instance, the radius function decreases .2 of a grid unit each time. The learning rate is a exponential decay function along with the intensity of influence of the matching sample within the radius.

One encountered difficulty is creating a visually appealing from the selected data. The fields mapped to the red and blue channels was greater in magnitude than the one in the green channel which created a overall color scheme that went from blue to red. The magnitude of the field mapped to the green channel is due to this type of data and not in relation to the magnitude of the red or blue channel data. An additional scaling factor to the green channel resulted in the user being able to perceive the influence of this field and a more appealing visualization.

Another difficulty was understanding the Processing framework display paradigm. An early implementation that worked within a giant for-loop was not properly interacting and redrawing the screen. Once the code was converted to use Processing’s setup() function and automatically repeating draw() function to draw each “frame” of the output, the difficulty of implementing the SOM in Processing decreased.

A continuing limitation is the application of adding increased dimensions to the visualization. Multiple SOM discussions use color channel as the means to visualize each field with the sample. It would be interesting to extend the visualization to more dimensions that would be allowed with just using the three color channels.

The wikipedia article for SOM was the initial source for understanding the technique. An article written by Tom Germano at (<http://davis.wpi.edu/~matt/courses/soms/>) described indivdual steps and provided ideas for algorithms that would sastify the concept of SOM. Another article titled Kohonen’s Self Organizing Feature Maps (<http://www.ai-junkie.com/ann/som/som1.html>) provided additional insight into the underlying concepts.