SOM visualizations

Density-based visualizations

- Hit Histogramm: coloring by number of vectors per unit
- Smoothed Data Histograms:
  - Each input vector is mapped not only onto best-matching unit, but onto n-best matching units
  - Creates smoothing effect
  - 4 methods:
    - counting all n units equally
    - weigth depends on distance: 1/d
    - normalized: 1/n: for 1. unit 1, then 1/2, 1/3, ...
    - normalized distance weight: 1/dn (dn : min-max normalized distance)
- P-Matrix:
  - Hypersphere around each weigth vector in data space
  - Counting the number of input vectors in this hypersphere
  - Good for emergent SOMs
  - shows density of data on units
- Sky Metaphor:
  - display them on their “exact” position within the unit, not just in the centre
  - Apply pull force to determine exact position (triangulation)
- Neighborhood Graphs:
  - which areas of the map are close to each other based on density in input space
  - reveals topology violations and cluster densities
  - Similar to P-matrix, but for pairs of units
  - Radius method:
    - Connect data vectors within a radius in the input space
    - Focus on density
  - KNN method:
    - Connect n nearest neighbors in input space
    - Focus on cluster cardinality

Distance-based visualizations

- Activity Histograms:
  - visualizes the distance between input vector and all weight vectors
  - per-data-point visualization
  - reveal cluster homogeneity/ topology violations
- Minimum Spanning Trees:
  - Either of input data or of weight vectors
  - Projected onto SOM
  - Not extremely useful in most cases
- Cluster Connections (CC)
  - Calculate distance between neighboring units
  - Draw connecting lines between units according to thresholds

- D-Matrix, U-Matrix:
  - D-Matrix:
    - average distance to all neigboring units
    - rarely used any more
  - U-Matrix:
    - distance to each neighboring units, interpolate
    - can reveal cluster boundaries
- U* Matrix: combination of U-Matrix + P-Matrix
- Vectorfields: Flow / Borderline:
  - Show which areas of the map are close to each other
  - similar to U-matrix, but for pairs of units
  - Flow
    - arrows point to cluster centers
    - length shows intensity
  - Borderlines
    - dual representation
    - rotate arrows by 90 degrees
    - show cluster boundaries

Class Info visualizations

- Pie Charts / Patches
- Class Coloring:
  - Chessboard:
    - Voronoi Tesselation
    - color voronoi cells with patches accoridng to percentual share of class
  - Color Filling with Attractor:
    - Voronoi Tesselation
    - Fill with dominant class color
    - Identify neighboring class distributions
    - Identify attractors
    - Flood-fill style coloring along attractors according to frequency
  - For both there is an optional class threshold value

Attribute-based visualizations

- Component Planes:
  - One visualization per attribute
  - color each unit according to the value of a given attribute
  - Allows for analysing regularity of distribution
  - Can be hard to understand correlations between attributes
- Clustering of Component Planes
- Metro Maps:
  - aggregation of correlated component planes
  - Steps:
    - Discretization of component planes
    - component lines (revealing gradients)
    - Aggregation
    - Selection (little to no scattering /local minima)
    - Snapping
- Vectorfields: grouped Flow (for groups of attributes)

Clusterings

- flat: k-means
- hierarchical:
  - clusters are merged to form higher-level aggregation
  - single/complete linkage
  - WARD's linkage