Activity Mathematics:
Pattern Classification in Dense Sensor Fields
A brief summary of work to date

James Howard
Colorado School of Mines

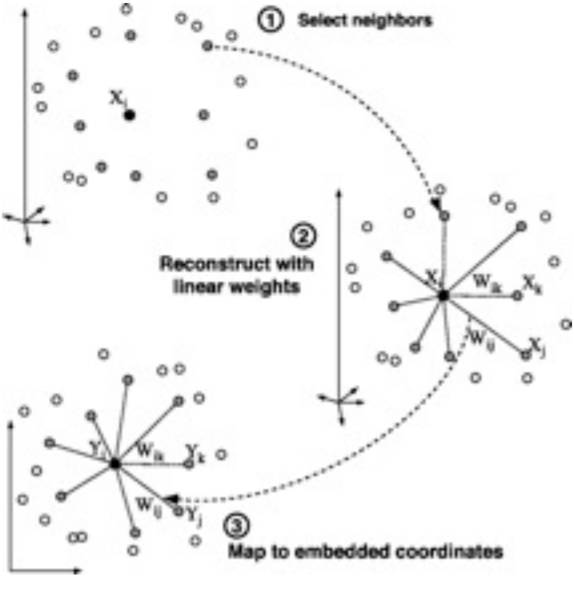
Major Ideas

- Dimensionality Reduction
- Clustering
- Pattern Matching
- Other Ideas

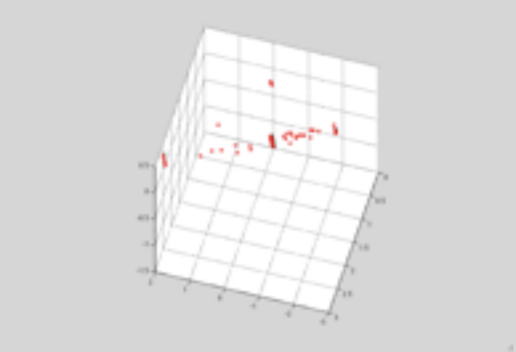
Major Ideas

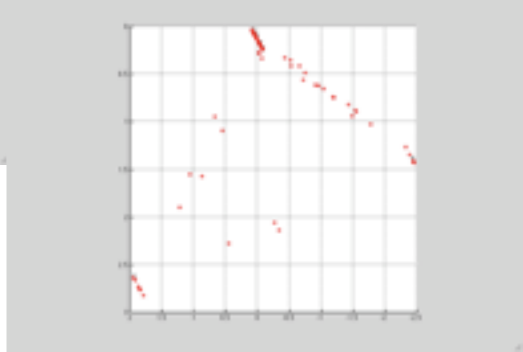
- Dimensionality Reduction
- Clustering
- Pattern Matching
- Other Ideas

Locally Linear Embedding



LLE - Results





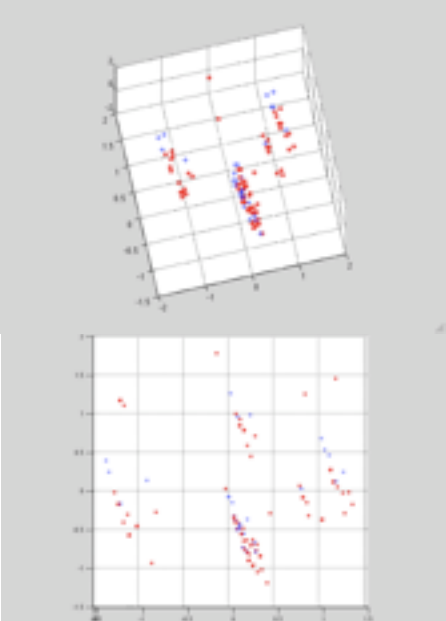
```
Warning: Matrix is close to singular or badly scaled.
         Results may be inaccurate. RCOND = 3.700743e-17.
> In eigs>AminusSigmaBsolve at 1204
  In eigs at 257
  In IleMod at 119
  In run at 69

Warning: Matrix is close to singular or badly scaled.
         Results may be inaccurate. RCOND = 3.700743e-17.
> In eigs>AminusSigmaBsolve at 1204
  In eigs at 257
  In IleMod at 119
  In run at 69
Done.
```

LLE - What's left?

- Assumption of locally linear data valid?
- AMPLE
- Distance metric definition
  - Walk and Loiter distance

Principal Component Analysis

$$S = \begin{matrix} 5.6482 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5.4354 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4.6176 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4.5500 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3.9379 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3.5706 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2.5820 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.3930 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.2864 \end{matrix}$$



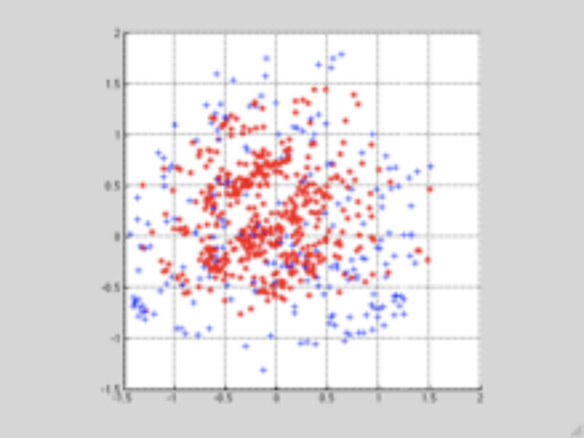
Dataset: (Dense)

1000 Points

100 Loitering Events

100 Walk Left Events

100 Walk Right Events



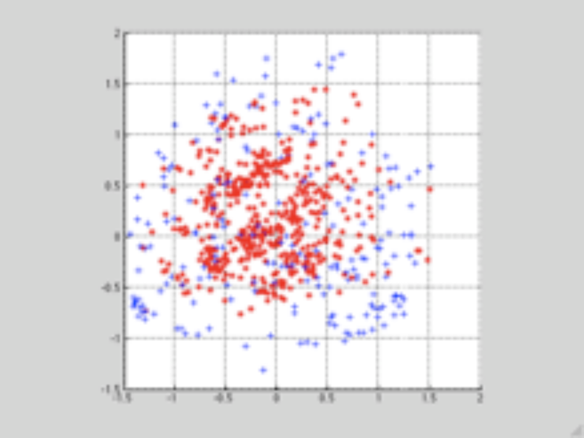
Dataset: (Sparse)

1000 Points

10 Loitering Events

10 Walk Left Events

10 Walk Right Events



Major Ideas

- Dimensionality Reduction
- Clustering
- Pattern Matching
- Other Ideas

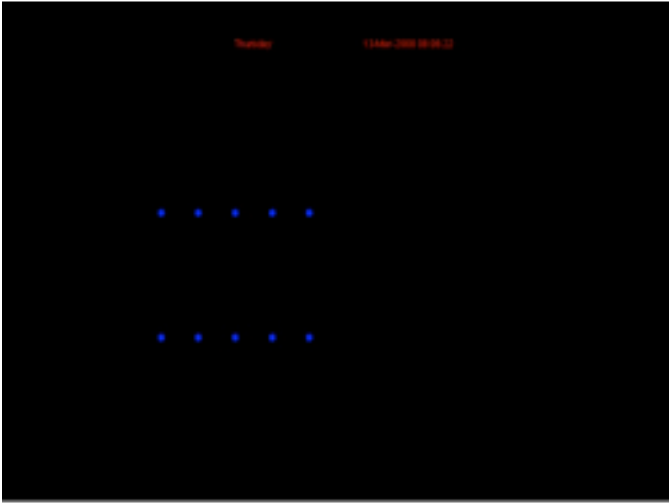
Closest Pattern Matching (CPM)

	0	0	0	0	0	0
	1	0	0	0	0	0
	0	1	1	0	1	1
	1	0	0	1	0	0
	1	0	0	1	1	1
	0	1	1	0	0	1
Predicted Row	0	1	1	0	0	0

CPM - Continued

	0	0	0	0	0	0
	1	0	0	0	0	0
	0	1	1	0	1	1
	1	0	0	1	0	0
	1	0	0	1	1	1
	0	1	1	0	0	1
Predicted Row	0	1	1	0	0	0

CPM - Continued



Growing Pattern CPM

0	0	0	0	0	0
1	0	0	0	0	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	0	1	1	1
0	1	1	0	0	1
0	1	1	0	0	0

0	0	0	0	0	0
1	0	0	0	0	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	0	1	1	1
0	1	1	0	0	1
0	1	1	0	0	0

0	0	0	0	0	0
1	0	0	0	0	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	0	1	1	1
0	1	1	0	0	1
0	1	1	0	0	0

0	0	0	0	0	0
1	0	0	0	0	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	0	1	1	1
0	1	1	0	0	1
0	1	1	0	0	0

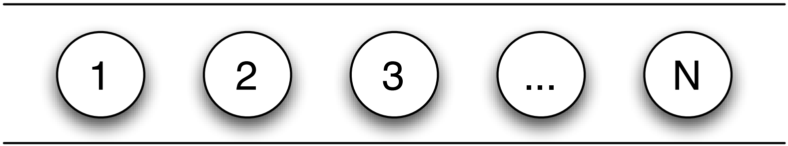
0	0	0	0	0	0
1	0	0	0	0	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	0	1	1	1
0	1	1	0	0	1
0	1	1	0	0	0

Growing Pattern CPM

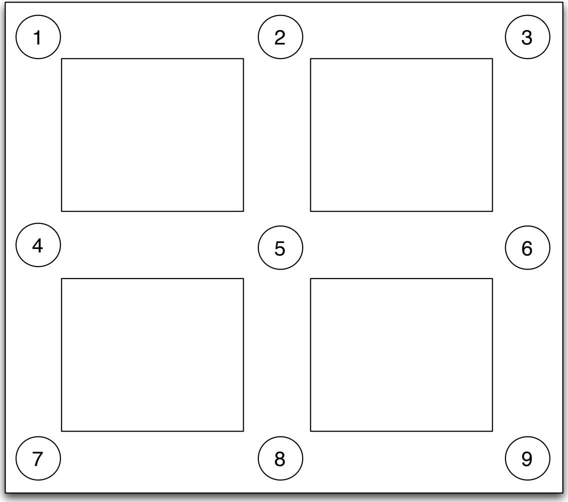
0	0	0	0	0	0
1	0	0	0	0	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	0	1	1	1
0	1	1	0	0	1
0	1	1	0	0	0

Growing Pattern CPM

CPM Problem

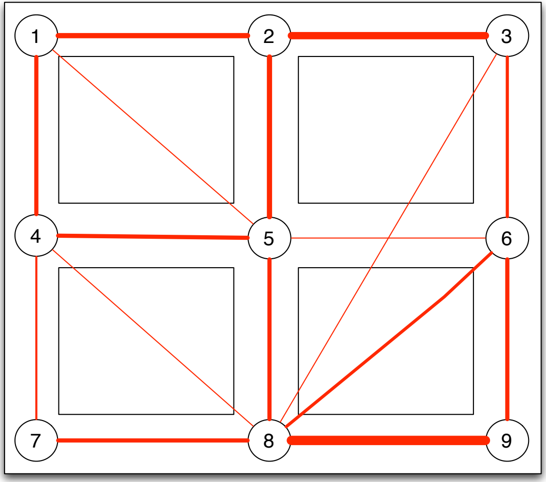


CPM - Problem



CPM - Future

Dynamically create patterns based on correlation score.



CPM - Future

Pearson Correlation

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\,\sqrt{E(Y^2) - E^2(Y)}}.$$

Entropy Correlation

$$H(X) = \mathrm{E}(I(X)) \;=\; -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$$

$$\begin{aligned} H(Y|X) &\stackrel{\text{def}}{=} \sum_{x\in\mathcal{X}} p(x)\, H(Y|X=x) \\ &= -\sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} p(y|x)\,\log\, p(y|x) \\ &= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(y,x)\,\log\, p(y|x) \\ &= -E_{p(x,y)} \log\, p(y|x). \end{aligned}$$

CPM - Future

- Construct a model per sensor
- Preprocess the data
    - Compress over time
- Incorporate contextual models

Major Ideas

- Dimensionality Reduction
- Clustering
- Pattern Matching
- Other Ideas

Context Discovery and Model Linking

- How can we dynamically discover contexts using only state information?
  - Prior definition
  - Use a technique to dynamically create them.
    - Simulated Annealing with Latent Semantic Analysis as a move approximation metric
- Given a set of contexts do we need to aggregate them for better accuracy?

Non Pursued Paths

- Modifying CPM's pattern shape
- Match previously defined activities to events in dataset
- Attempt to classify nodes dynamically if they exhibit behavior outside the defined realm.
- Apply a boosting approach on top of a simple classifier