Using Cellular Automata as a clustering tool

James Hurford

13/10/2010

Introduction

- Data mining (Classification)
- Purpose is to show that cellular automata can be used as a clustering tool

Clustering and Classification tool

- Emergent behaviour
- Less error prone
- Low-bias and Self organising

The approach

- Uses a majority voting system
- Grid size and dimensions determined by predicates used
- It is run until convergence happens
  - This is achieved when all cells are assigned a class

My implementation

- Written in C++ using the STL
- The process
  - Define a grid
    - Predicates must be turned into integers
  - Populate with training data
  - Run until finished
  - Test with test data

Findings

- In general my data does seem to support Fawcett's hypothesis
- Does not exactly match up with Fawcett's results

Explanation/Evaluation

- Inconsistencies between my results and Fawcett's
  - Maybe because of a lack of specific information
- The process can take a long time
- Memory is the bottleneck

Conclusions

- Cellular automata can be used as a clustering and classification tool