MCMC Span Experiment Report

LI Tao

October 8, 2012

Introductin

- Problem: Mining frequent pattern in large-scale database efficiently.
- Basic Idea: Use MCMC randomly traverse the data, and retrieve proper sequence fulfil the requirement.

Algorithm Overview

1. Starts from null state as the current state
2. If current state is max pattern, store it to the dictionary, and revert the current state to null.
3. Calculate all the next state of the current state.
4. Check the frequency of each next state, erase those lower than min-support.
5. Randomly pick one next state as the current state.
6. Repeat steps 2 to 5 until the size of dictionary is 20.

Experiment Detail

- Experiment Data:
  - Trajectory of handwriting (normalized and discretized),
  - The character of the handwriting as labels, **8** labels are selected.
  - Data Size:1236
- Experiment:
  - Configuration: min-support of the frequent pattern: 100
  - Accuracy of svm-prediction: 64.5833%
- Comparison: In the same configuration, prefixspan got 59.25% accuracy.