CS 147:
Computer Systems Performance Analysis
Workload Characterization

CS 147:
Computer Systems Performance Analysis
Workload Characterization

Overview

Terminology

Specifying Parameters

Identifying Parameters
- Histograms
- Principal-Component Analysis
- Markov Models

Clustering
- Clustering Steps
- Clustering Methods
- Using Clustering

Overview

Terminology

Specifying Parameters

Identifying Parameters
Histograms
Principal-Component Analysis
Markov Models

Clustering
Clustering Steps
Clustering Methods
Using Clustering

Workload Characterization Terminology

- *User* (maybe nonhuman) requests service
  - Also called *workload component* or *workload unit*
- *Workload parameters* or *workload features* model or characterize the workload

Workload Characterization Terminology

- *User* (maybe nonhuman) requests service
  - Also called *workload component* or *workload unit*
- *Workload parameters* or *workload features* model or characterize the workload

Selecting Workload Components

- Most important: components should be *external*: at interface of SUT
- Components should be homogeneous
- Should characterize activities of interest to the study

Selecting Workload Components

- Most important: components should be *external*: at interface of SUT
- Components should be homogeneous
- Should characterize activities of interest to the study

Choosing Workload Parameters

- Select parameters that depend only on workload (not on SUT)
- Prefer controllable parameters
- Omit parameters that have no effect on system, even if important in real world

Choosing Workload Parameters

- Select parameters that depend only on workload (not on SUT)
- Prefer controllable parameters
- Omit parameters that have no effect on system, even if important in real world

Averaging

- Basic character of a parameter is its average value
- Not just arithmetic mean
- Good for uniform distributions or gross studies

Averaging

- Basic character of a parameter is its average value
- Not just arithmetic mean
- Good for uniform distributions or gross studies

Specifying Dispersion

- Most parameters are non-uniform
- Specifying variance or standard deviation brings major improvement over average
- Average and s.d. (or C.O.V.) together allow workloads to be grouped into classes
  - Still ignores exact distribution

Specifying Dispersion
- Most parameters are non-uniform
- Specifying variance or standard deviation brings major improvement over average
- Average and s.d. (or C.O.V.) together allow workloads to be grouped into classes
  - Still ignores exact distribution

Single-Parameter Histograms

- Make histogram or kernel density estimate
- Fit probability distribution to shape of histogram
- Chapter 27 (not covered in course) lists many useful shapes
- Ignores multiple-parameter correlations

Single-Parameter Histograms

- Make histogram or kernel density estimate
- Fit probability distribution to shape of histogram
- Chapter 27 (not covered in course) lists many useful shapes
- Ignores multiple-parameter correlations

Multi-Parameter Histograms

- Use 3-D plotting package to show 2 parameters
  - Or plot each datum as 2-D point and look for “black spots”
- Shows correlations
  - Allows identification of important parameters
- Not practical for 3 or more parameters

Multi-Parameter Histograms

- Use 3-D plotting package to show 2 parameters
  - Or plot each datum as 2-D point and look for “black spots”
- Shows correlations
  - Allows identification of important parameters
- Not practical for 3 or more parameters

Principal-Component Analysis (PCA)

- How to analyze more than 2 parameters?
- Could plot endless pairs
  - Still might not show complex relationships
- Principal-component analysis solves problem mathematically
  - Rotates parameter set to align with axes
  - Sorts axes by importance

Principal-Component Analysis (PCA)

- How to analyze more than 2 parameters?
- Could plot endless pairs
  - Still might not show complex relationships
- Principal-component analysis solves problem mathematically
  - Rotates parameter set to align with axes
  - Sorts axes by importance

Advantages of PCA

- Handles more than two parameters
- Insensitive to scale of original data
- Detects dispersion
- Combines correlated parameters into single variable
- Identifies variables by importance

Advantages of PCA

- Handles more than two parameters
- Insensitive to scale of original data
- Detects dispersion
- Combines correlated parameters into single variable
- Identifies variables by importance

Disadvantages of PCA

- Tedious computation (if no software)
- Still requires hand analysis of final plotted results
- Often difficult to relate results back to original parameters

Disadvantages of PCA

- Tedious computation (if no software)
- Still requires hand analysis of final plotted results
- Often difficult to relate results back to original parameters

Markov Models

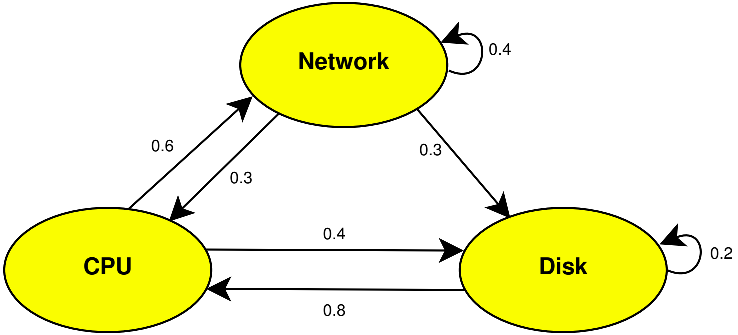
- Sometimes, distribution isn't enough
- Requests come in sequences
- Sequencing affects performance
- Example: disk bottleneck
  - Suppose jobs need 1 disk access per CPU slice
  - CPU slice is much faster than disk
  - Strict alternation uses CPU better
  - Long disk-access strings slow system

Markov Models

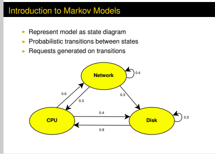
- Sometimes, distribution isn't enough
- Requests come in sequences
- Sequencing affects performance
- Example: disk bottleneck
  - Suppose jobs need 1 disk access per CPU slice
  - CPU slice is much faster than disk
  - Strict alternation uses CPU better
  - Long disk-access strings slow system

Introduction to Markov Models

- Represent model as state diagram
- Probabilistic transitions between states
- Requests generated on transitions



2015-06-15


Introduction to Markov Models
- Represent model as state diagram
- Probabilistic transitions between states
- Requests generated on transitions

Creating a Markov Model

- Observe long string of activity
- Use matrix to count pairs of states
- Normalize rows to sum to 1.0

	CPU	Network	Disk
CPU		0.6	0.4
Network	0.3	0.4	0.3
Disk	0.8		0.2

Creating a Markov Model
- Observe long string of activity
- Use matrix to count pairs of states
- Normalize rows to sum to 1.0

	CPU	Network	Disk
CPU		0.6	0.4
Network	0.3	0.4	0.3
Disk	0.8		0.2

Example Markov Model

- Reference string of opens, reads, closes: ORORRCOORCRRRRCC
- Pairwise frequency matrix:

	Open	Read	Close	Sum
Open	1	3		4
Read	1	4	3	8
Close	1	1	1	3

Example Markov Model

- Reference string of opens, reads, closes: ORORRCOORCRRRRCC
- Pairwise frequency matrix:

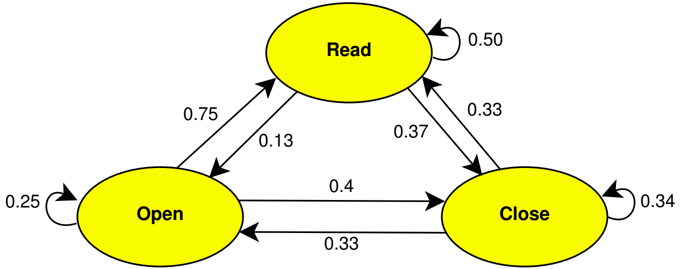
	Open	Read	Close	Sum
Open	1	3		4
Read	1	4	3	8
Close	1	1	1	3

Markov Model for I/O String

- Divide each row by its sum to get transition matrix:

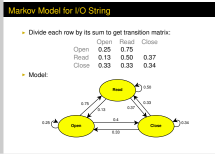
	Open	Read	Close
Open	0.25	0.75	
Read	0.13	0.50	0.37
Close	0.33	0.33	0.34

- Model:



2015-06-15

CS147
Identifying Parameters
Markov Models
Markov Model for I/O String



Clustering

- Often useful to break workload into categories
- "Canonical example" of each category can be used to represent all samples
- If many samples, generating categories is difficult
- Solution: clustering algorithms

Clustering

- Often useful to break workload into categories
- "Canonical example" of each category can be used to represent all samples
- If many samples, generating categories is difficult
- Solution: clustering algorithms

Steps in Clustering

- Select sample
- Choose and transform parameters
- Drop outliers
- Scale observations
- Choose distance measure
- Do clustering
- Use results to adjust parameters, repeat
- Choose representative components

Steps in Clustering

- Select sample
- Choose and transform parameters
- Drop outliers
- Scale observations
- Choose distance measure
- Do clustering
- Use results to adjust parameters, repeat
- Choose representative components

Selecting A Sample

- Clustering algorithms are often slow
  - Must use subset of all observations
- Can test sample after clustering: does every observation fit into some cluster?
- Sampling options
  - Random
  - Heaviest users of component under study

Selecting A Sample

- Clustering algorithms are often slow
  - Must use subset of all observations
- Can test sample after clustering: does every observation fit into some cluster?
- Sampling options
  - Random
  - Heaviest users of component under study

Choosing and Transforming Parameters

- Goal is to limit complexity of problem
- Concentrate on parameters with high impact, high variance
  - Use principal-component analysis
  - Drop a parameter, re-cluster, see if different
- Consider transformations such as Sec. 15.4 (logarithms, etc.)

Choosing and Transforming Parameters

- Goal is to limit complexity of problem
- Concentrate on parameters with high impact, high variance
  - Use principal-component analysis
  - Drop a parameter, re-cluster, see if different
- Consider transformations such as Sec. 15.4 (logarithms, etc.)

Dropping Outliers

- Must get rid of observations that would skew results
  - Need great judgment here
  - No firm guidelines
- Drop things that you know are “unusual”
- Keep things that consume major resources
  - E.g., daily backups

Dropping Outliers

- Must get rid of observations that would skew results
  - Need great judgment here
  - No firm guidelines
- Drop things that you know are “unusual”
- Keep things that consume major resources
  - E.g., daily backups

Scaling Observations

- Cluster analysis is often sensitive to parameter ranges, so scaling affects results
- Options:
  - Scale to zero mean and unit variance
  - Weight based on importance or variance
  - Normalize range to [0, 1]
  - Normalize 95% of data to [0, 1]

2015-06-15

Scaling Observations

- Cluster analysis is often sensitive to parameter ranges, so scaling affects results
- Options:
  - Scale to zero mean and unit variance
  - Weight based on importance or variance
  - Normalize range to [0, 1]
  - Normalize 95% of data to [0, 1]

Choosing a Distance Measure

- Endless possibilities available
- Represent observations as vectors in k-space
- Popular measures include:
  - Euclidean distance, weighted or unweighted
  - Chi-square distance
  - Rectangular (“Manhattan”) distance

Choosing a Distance Measure

- Endless possibilities available
- Represent observations as vectors in k-space
- Popular measures include:
  - Euclidean distance, weighted or unweighted
  - Chi-square distance
  - Rectangular (“Manhattan”) distance

Chi-square distance is:

$$d = \sum_{k=1}^{n} \left\{ \frac{(x_{ik} - x_{jk})^2}{x_{ik}} \right\}$$

and requires $x_{\cdot k}$ to be close together or low values of $x_{\cdot k}$ will over-weight parameters. Used primarily in distribution fitting.

Clustering Methods

- Many algorithms available
- Computationally expensive (NP to find optimum)
- Can be simple or hierarchical
- Many require you to specify number of desired clusters
- Minimum Spanning Tree (from book) is not only option!

Clustering Methods

- Many algorithms available
- Computationally expensive (NP to find optimum)
- Can be simple or hierarchical
- Many require you to specify number of desired clusters
- Minimum Spanning Tree (from book) is not only option!

Types of Clustering

- Agglomerative vs. divisive
- Hierarchical vs. non-hierarchical

Types of Clustering

- Agglomerative vs. divisive
- Hierarchical vs. non-hierarchical

Minimum Spanning Tree Clustering

- Start with each point in a cluster
- Repeat until single cluster:
  - Compute centroid of each cluster
  - Compute intercluster (inter-centroid) distances
  - Find smallest distance
  - Merge clusters with smallest distance
- Result is a hierarchy of clusters
- Method produces stable results
  - But not necessarily optimum

Minimum Spanning Tree Clustering

- Start with each point in a cluster
- Repeat until single cluster:
  - Compute centroid of each cluster
  - Compute intercluster (inter-centroid) distances
  - Find smallest distance
  - Merge clusters with smallest distance
- Result is a hierarchy of clusters
- Method produces stable results
  - But not necessarily optimum

K-Means Clustering

- One of most popular methods
- Number of clusters is input parameter, k
- First randomly assign points to clusters
- Repeat until no change:
  - Calculate center of each cluster: $(\overline{x}, \overline{y})$
  - Assign each point to cluster with nearest center
- Big problem: How to choose k
  - Prior knowledge
  - Trial and error

K-Means Clustering

- One of most popular methods
- Number of clusters is input parameter, k
- First randomly assign points to clusters
- Repeat until no change:
  - Calculate center of each cluster: $(\overline{x}, \overline{y})$
  - Assign each point to cluster with nearest center
- Big problem: How to choose k
  - Prior knowledge
  - Trial and error

Jarvis & Patrick's Method

- Start with each point in own cluster
- For each point, make list of *n* closest other points
- For each point pair, if *k* of *n* nearest neighbors are shared, combine their clusters
- Finds non-globular clusters
- Extremely sensitive, in non-intuitive ways, to *k* and *n*

Jarvis & Patrick's Method

- Start with each point in own cluster
- For each point, make list of *n* closest other points
- For each point pair, if *k* of *n* nearest neighbors are shared, combine their clusters
- Finds non-globular clusters
- Extremely sensitive, in non-intuitive ways, to *k* and *n*

Interpreting Clusters

- Art, not science
- Drop small clusters (if little impact on performance)
- Try to find meaningful characterizations
- Choose representative components
  - Number proportional to cluster size or to total resource demands

Interpreting Clusters

- Art, not science
- Drop small clusters (if little impact on performance)
- Try to find meaningful characterizations
- Choose representative components
  - Number proportional to cluster size or to total resource demands

Drawbacks of Clustering

- Clustering is basically AI problem
- Humans will often see patterns where computer sees none
- Result is extremely sensitive to:
  - Choice of algorithm
  - Parameters of algorithm
  - Minor variations in points clustered
- Results may not have functional meaning

Drawbacks of Clustering

- Clustering is basically AI problem
- Humans will often see patterns where computer sees none
- Result is extremely sensitive to:
  - Choice of algorithm
  - Parameters of algorithm
  - Minor variations in points clustered
- Results may not have functional meaning