E-Center NPC Forecasting

An e-Center goal is to forecast network performance characteristics (NPCs), including channel utilization, delay, and packet loss.

Different candidate methodologies were tested for forecasting NPCs.

Three candidate methodologies were considered in detail:

- Neural networks
- Box-Jenkins ARIMA modeling
- Gaussian process regression (GPR)

GPR is the methodological framework we have selected.

Michael Frey
Bucknell University
10/26/2010

Gaussian Process Regression

- Background -

Statistical research community (1970s):

> GPR assumes that the NPC being modeled is evolving along a sample path of a Gaussian process and selects the path that is most consistent (maximum likelihood) with the estimated mean and covariance of the process.

Machine learning community (1990s to present):

- Form of supervised learning
- Closely related to support vector machines
- Limiting form of an infinitely large neural network.

Michael Frey
Bucknell University
10/26/2010

Gaussian Process Regression

- Statistical Optimality -

GPR predictions are *unbiased*.

GPR predictions have *minimum* variance.

GPR readily *estimates* prediction standard errors.

- The statistical component of the e-Weather Center is specified to have a capability to predict NPCs with accompanying estimates of prediction standard error.

- These estimates of prediction standard error yield prediction intervals with any chosen declared confidence level.

Michael Frey
Bucknell University
10/26/2010

Gaussian Process Regression

- Periodic, synchronous data -

Periodic, synchronously collected NPC data can be collected on ESNet, but data with this structure are unrealistic on the broader Internet beyond ESNet.

E-Center forecasting must accommodate NPC data collected:

- Aperiodically
- Asynchronously on different network elements

GPR does not depend on any time structure within its “training” data; GPR accepts both periodically and aperiodically collected data and synchronous and asynchronous measurements on different network elements.

Michael Frey
Bucknell University
10/26/2010

Gaussian Process Regression

- Structural data, active measurements -

E-Center may collect structural data on, for example, correlations between successive hops to compensate for missing data.

E-Center may initiate active NPC measurements in response to missing data.

E-Center forecasting should be able to exploit:

- Structural data (e.g., on correlations between successive hops)
- Active measurements

GPR gracefully accepts both structural data (using it to estimate hyperparameters) and active measurement data (classifying it without distinction with other training data).

Michael Frey
Bucknell University
10/26/2010

Gaussian Process Regression

- NPC periodicity -

ESnet NPC data show little periodicity, but periodicities (diurnal and otherwise) are clearly evident in NPC data from the broader internet.

GPR readily adapts to periodicities in NPC data and exploits periodicity for forecasting.

Michael Frey
Bucknell University
10/26/2010

Gaussian Process Regression

- Scalability -

NPC data can be collected for every link and site-to-site path on ESNet. This is not scalable to the broader Internet.

E-Center's forecasting component must therefore eventually be capable of network tomography; i.e., it must have a robust capability for inference from and between various combinations of

1) site-to-site NPC measurements
2) link NPC measurements
3) route determinations

Network tomography is a *hard* problem with presently no good, mature, general solution.

GPR can *potentially* perform network tomographic inferences.

Michael Frey
Bucknell University
10/26/2010

Gaussian Process Regression

- Implementation -

MatLab/C software has been written to implement GPR.

This implementation exploits efficient, reliable *MatLab* routines and requires *MatLab* to be running during execution.

This software was delivered July 8, 2010 with documentation and is being installed.

Michael Frey
Bucknell University
10/26/2010

E-Center Statistics Architecture

- Anomaly-centric -

An anomaly is any occurrence or object that is strange, unusual, or unique—a discrepancy or deviation from an established pattern or trend.

Anomalies are value-neutral.

An anomaly's significance—good/bad—is interpreted within the context of the particular user-specific application.

Michael Frey
Bucknell University
10/26/2010

E-Center Statistics Architecture

- Layered -

Layer	Input	Process(es)	Product(s)
Layer 3: Interpretation layer	Layer 2 anomalies	User-specific applications - expert system-based - data fusion (e.g., input from network manager logs)	User-specific information on (e.g.): - network status - resource availability - quality-of-service - fault diagnosis - DDoS attack alarm
Layer 2: Anomaly layer	Layer 1 data	Statistical process charting (Process control charts)	Anomaly - detection - classification - presentation
Layer 1: Data layer	ESnet	perfSONAR, pingER, traceroute, other	Data management - collection - caching - presentation, forecasting

Michael Frey
Bucknell University
10/26/2010

E-Center Statistics Architecture

- Advantages -

Standard tools available for anomaly detection and classification:

- Statistical process control charts

Enables fusion with other kinds of data

- Anomaly data are qualitative, NPC data are quantitative.

Data reduction

- Anomaly data are relatively much less capacious than are NPC data.

Michael Frey
Bucknell University
10/26/2010

E-Center Statistics Architecture

- Statistical process control charts -

- Allow *simple* detection of events (anomalies) that indicate process change (of whatever sort).
- Anomalies can be difficult to detect when NPC is continuously varying; control charts offer *statistically objective* change criteria.
- Different charts identify different types of anomaly, allowing simple anomaly *classification*.
- Individual NPCs and *vectors* of NPCs can be charted; e.g., packet loss on different hops of a network path.
- Control charts time-aggregate NPCs – visualization on *different time-scales*.

Michael Frey
Bucknell University
10/26/2010

E-Center

- Research in Network Statistics -

Augment E-Center to let users download visualized data directly from e-Center interface.

- Write instruction manual for a graduate course based on e-Center.
- Promote availability of e-Center at statistics meetings (annual JSM).
- Announce e-Center to university statistics departments.

Michael Frey
Bucknell University
10/26/2010