数理システム知識工学部の御紹介

知識工学部

Mathematical Systems Inc.

2012-11-22

グラフ時系列によるネットワーク異常検出

概要

1. ネットワークノード集合をグラフとして捉え、そのトポロジーの崩れをノード相互作用における異常とみなす方法。
2. network traffic や 其の他のデータを使って検証し、operator の気付かない異常を発見するなど、それなりの効果を出してきている。
3. オンライン性に優れ、大規模ネットワークに対応できる。

特典

- ネットワークに限らず機器故障診断など、多変量時系列解析一般に適用できる。
- 車載計測器データからの異常解析にも、類似手法が適用されている。

グラフ時系列によるネットワーク異常検出 cont.

以下のような理論があり、それぞれを自前で実装し、解析に利用している。

実装

1. Ide-Kashima (Eigenspace-based Anomaly Detection in Computer Systems)
2. SNN (Computing Correlation Anomaly Scores using Stochastic Nearest Neighbors)
3. EEC (Network anomaly Detection based on Eigen Equation Compression)

Time Seriese

先の手法は、多くのノードの時系列データをまとめて見るものであるが、時系列一般についても、代表的なモデルは全て自作し、自分達のノウハウにしてきている。(デモサイトあり)

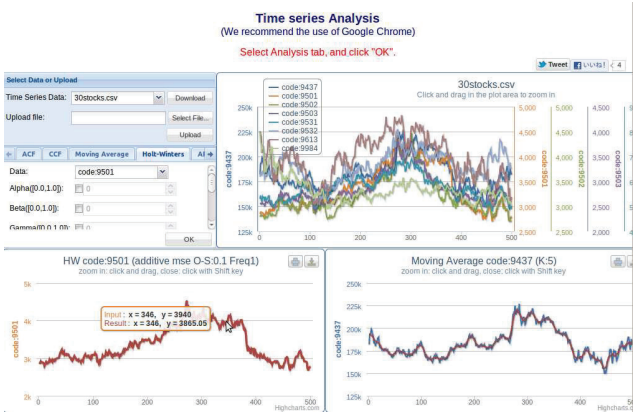
Holt-winters モデル

- 時系列データの水準・トレンド・周期性の各成分を指数平滑化法によって表現したモデル
- パラメータの自動推定

変化点検出 (ChangeFinder)

- 外れ値検出と異なり、時系列の振る舞いの変化を捉える
- SDAR モデル (オンライン忘却型時系列モデル)

Time Seriese demo (http://cl-www.msi.co.jp:8100/time-series/index.html)



機械学習

分析器は、全て Common Lisp で書かれた機械学習パッケージ上で実現。後で示す国立情報学研究所プロジェクトや、2lslnc での利用実績あり。

CLML

- 一つの実行形式で動作し、種々のライブラリーに依存しないためインストールの手間が不要。(クラウド向き)
- 動作プラットフォームを選ばない (Windows32/64, Linux32/64, Solaris, MacOS)
- 並列実行、分散処理に適する開発基盤を持つ (SMP, forkfuture)

High Performance Network Tools

数十万規模のターゲットを想定した高速なネットワークツールを開発した。OS の kernel まで踏込むプログラミング。

CL-Ping

Interleave retry mode, ICMP or UDP, Distributed, IPv4 & IPv6

CL-Traceroute

Traceroute to multiple targets, Path merge, Simultaneous search

CL-SNMP

Multi-target, multi-var GET and GETNEXT, and SNMP walk. Packet rate throttling.

Routing Protocol

種々のルーティングプロトコルに根差したアプリケイション開発経験が豊富。

ENCORE (http://www.ntt.co.jp/news/news01/0108/010830.html)

経路 hijack 監視 (BGP-4)

Topology Viewer & Network Debugger

BGP パケットから Internet の地図を書く (BGP-4) ネットワークで何か起きた時の解析ツールとして、ネットワークデバッガ的なものを目指している

IMS & Crawler

次世代ネットワークの根幹技術である、IP multimedia subsystem を開発した。

高可用性 IMS

- OpenIMS の可用性を 99.98% まで向上
- 徹底的試験とデバッグ

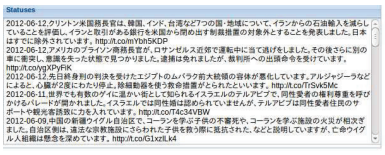
Crawler

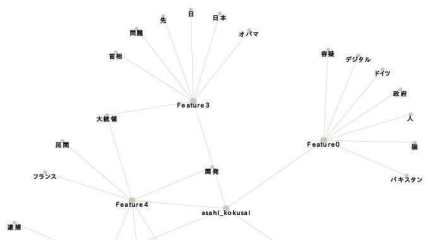
特別なライセンスを必要としない、高機能 twitter crawler を開発した。デモサイトあり。

High Performance Twitter Crawler

- 12.0 million twits per day
- 特別な license を必要としない

Twitter Demo Page (http://cl-www.msi.co.jp:8100/twitter-nmf.html)





リバースエンジニアリング

仕様書を自動生成

レガシー化したプログラムソースから、仕様書を自動生成し、保守性の向上に寄与するプロジェクト。

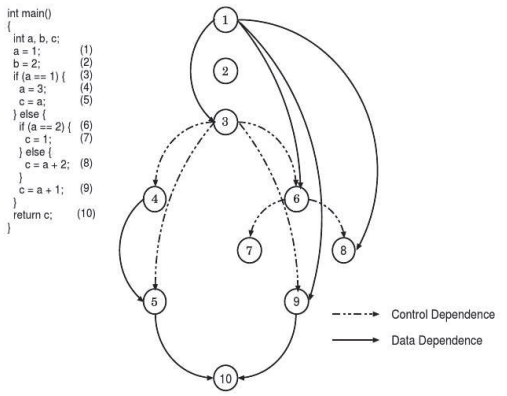
セキュリティ診断

セキュリティ診断 (SQL Injection の検知) プロジェクトでも同じ技術を適用。

コンパイラー技術

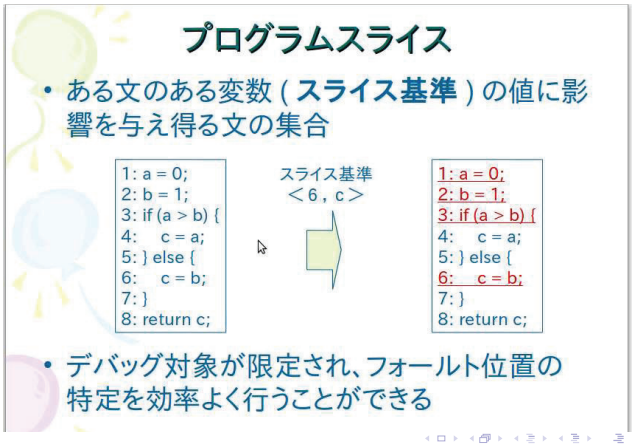
数理システムが独自に持つコンパイラー技術を驅使している。

プログラム依存グラフ (PDG)

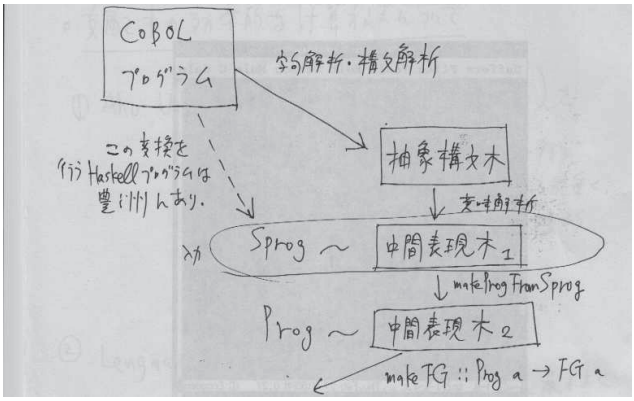


プログラムスライス

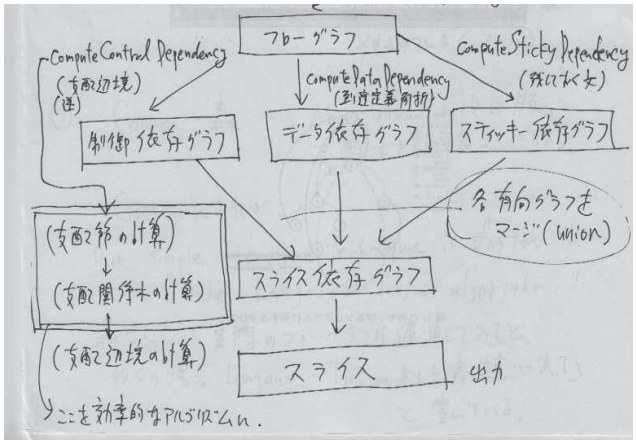
http://sel.ist.osaka-u.ac.jp/lab-db/Bthesis/archive/94/94.ppt



プログラムの解析



依存グラフを効率的なアルゴリズムで求める



Name Disambiguation

機械学習 (CLML)

プラットフォーム非依存な並列機械学習パッケージを使って、著者と論文の名寄せ (同定) を行なう。ci.nii.ac.jp のバックエンドで実稼働中。

全文検索エンジン

代表的全文検索エンジンとして知られる Lucene の 100 倍の性能を持つものを自作して適用。

並列 & 大容量

64 core, 5TB memory の機械で、一億件の書誌データを処理。

Name Disambiguation (http://ci.nii.ac.jp/)

Semantic 技術

Ontology, Description Logic, RDF, SPARQL, OWL といった、Semantic 技術を応用した開発経験を持つ。

1. 東北大学介護データ分析
2. 人事データ解析
3. 個人情報管理への適用

Semantic 技術の基本要素

RDF (Resource Description Framework)

グラフデータベース (データはネットワークである)

SPARQL (SPARQL Protocol and RDF Query Language)

論理に基いた共通の問合せ言語

SWRL (Semantic Web Rule Language)

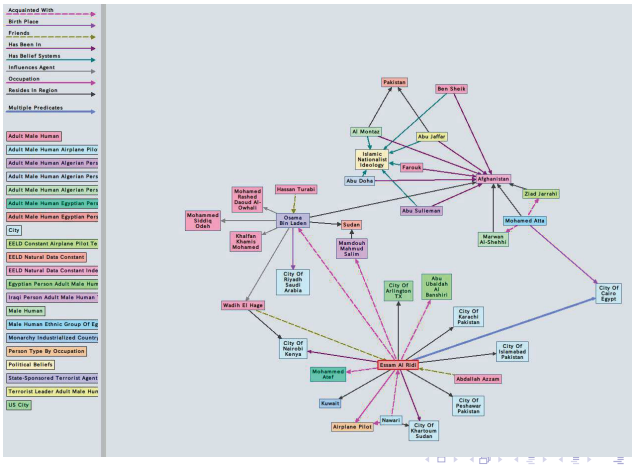
ルール記述による独自語彙の拡張

グラフデータベース

データマイニングをデータからの知識抽出と捉えると、抽出された知識を再利用できる形に整理保存するのがグラフデータベース。

1. データの解析結果をグラフデータベースに入れるだけで格段に見易くなる
   - 抽出されたルール、クラスタ、パターン
   - 話題と単語の繋がり、話題と人の繋がりなど
2. 一旦データベースに入れば検索言語で検索ができる
   - 全ての人が共通の検索基盤を持つ事になる
   - SaaS ならぬ DaaS (Data as a Service)
3. ルール記述により高度な検索語彙が利用できる
   - 語彙の独自拡張を許す
   - ストリームデータへビジネスルールの当嵌め

グラフデータベース (テロリストデータベース)



問合せ言語

見るだけではなくてネットワークを手繰るような問合せが可能。

Example (問合せ例)

```
(select (?A ?B ?X ?E)
  (q- OsamaBinLaden influencesAgent ?Z)
  (q- ?Z friends ?B)
  (q- ?Z hasBeenIn ?D)
  (q- ?B hasBeenIn ?D)
  (q- ?A friends ?B)
  (q- ?A occupation ?X)
  (q- ?B occupation ?X)
  (q- ?A birthPlace ?E)
  (q- ?B birthPlace ?E))
```


ルール記述

Example (Good Payer Time)

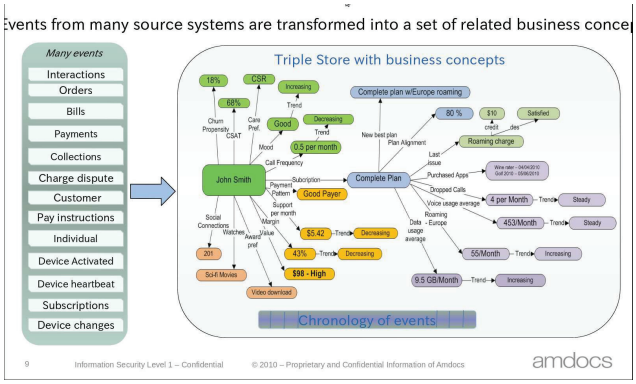
```
(<-- (good-payer-time ?n)
     (last-n-payments ?n ?avg ?sd)
     (< ?avg 14)
     (< ?sd 4))

(<-- (last-n-payments ?n ?avg ?sd)
     (last-n-bills ?n ?bills)
     (last-payments ?bills ?delays)
     (average/sd ?delays ?avg ?sd))

(<-- (last-payments ?bills ?delays) ...)

(<-- (average/sd ?list ?average ?sd) ...)
```

Semantic 技術の積極的応用例 (Telecom 顧客管理)



Semantic 技術がユーザー個別対応を 可能にする

up-to-date な顧客管理の流れ

1. Detect new event
   - 新たなイベントを検知し
2. Turn event into triples
   - イベントを RDF 表現
3. Recognize entity and call up all triples
   - イベントに関連するデータを識別し 抽出する
4. Apply business rules
   - ビジネスルールを当て嵌めて
5. Delete and add triples
   - データを適時更新する