JANOG21:
トラフィック流量の閾値監視に向けて

Appendix

- **本会議での質疑、コメント** (script**分**)
- **場外戦**: white board **のメモより**
- **発表者が得た知見、感想等**

February 2008

インターネットマルチフィード(株)
樽井 行保 (Yukiyasu Tarui)

本会議での質疑、コメント (1/2)

- サンプルデータの波形説明
  - どこのデータかって聞いても良い？ → それはノーコメントで。
- お客様の異常トラフィックを通知するために
  - 継続時間というパラメータも考えられる。
  - 回復時、アラーム解除も通知を受けたい。
- ベイジアンフィルタを応用できないか
  - 学習時：サンプル信号を与える。
  - 異常値は取り除いたほうがいい。
  - 正常、異常は見方で変わる。
  - 正解のデータをみなの力で。
- インターバルをどう考えるか？
  - 周期間隔によって変動するので監視間隔を5，10分間隔で同時にやってみる等
- モデルが変化するイベントが多くて難しい
  - 異常波形がパターン化できた時には人の動きが変化してパターンが適用できない可能性もあり。
- カレンダーは持っておいたほうがいい

本会議での質疑、コメント (2/2)

- 電力系統の考え方に流用できない？
  - 全体を見て、観測することもできない？
  - いい落しどころがほしい。
- トラフィック量の増減は経済に依存
  - トラフィックパターンの原因分析（経済予測）をするとよい。
  - ジッタ・ディレイのことも考えてほしい。
    - 今はとっかかりとして bps に着目しているだけ。
    - 時系列データであれば、なんでも解析は可能。
    - 他のシステムへの転用もできるだろう。
- KHアルゴリズム：パラメータの自動化
  - 周期を観測する手法も検討はしている（高速フーリエ変換：FFTとかどう？）
  - TLDアルゴリズムも同様
- 異常値を検出し、他のデータとも比較するアプローチもできる。
  - トラフィック流量のみの観測後に詳細な調査は必要
  - 原因の複合性。金融工学にもヒントがありそう。
- STA/LTA法
  - 地震予測に用いられている方法。
  - 地面の揺れをプロットし、波形変化を見て地震を確認、震度が決まる。

場外戦: white board のメモより (1/2)

- TLD法でも定時観測(周期性)を合わせると、定期的な（急激な）変化にも対応できるのでは？
  - 定時でもけっこう変動がある
- 他の業界の状況等
  - 株では分散から見て（株価の）予測値が出せるらしい一旦高くなると、しばらくその値段が続く、という感じ
  - 電力の予測 (e.g. デンコちゃん)
    - たとえばイベントの時はどう予測するか（ex.甲子園中継時に電力使用量ピークに）
  - 地震でも（今回の発表と同様に）閾値監視を応用している
    - 地面の動き(=揺れ)をプロット。ノイズ成分を除去し、震度を計測。
    - ナイキストの定理 (サンプリング定理)
- 早朝帯（トラフィックが低い時間の変化）の動きを無視してよいのか？
  - そのトラフィックを見る場所によっても違う
    - ex. エッジはそこまで見ないが、コアでどこかで何かあった可能性があれば、通知してほしい
  - 他のデータとの組み合わせで分かることがあるかもしれない。（ワームの活動、DNSのquery、BGPのupdate、etc.）

場外戦: white board のメモより (2/2)

- 閾値は固定？動的？
  - 見る場所によっては固定で十分。
- 複合的なツール、システムを構築して正解率を上げる。
  - 誤検知(FP), 見逃し(FN) では、FN の方が罪。
  - ツールの冗長化 → ネットワークと一緒。
  - その際、ツール毎の特性を把握しておく必要がある。
    - 天気予報も複数の予測法を元に予報案を作り、日々、中の人が妥当性を判断して採用しているようだ。
- どこまでデータをtracebackさせるか？1週間？1カ月？1年？
  - 1年間 保持できると年間の季節変動まで見れるが。
- 検知した結果をfeedbackする仕組みは必要ではないか？
  - 使われていないがICMPのsource quenchなどを使うべき？

得られた知見、感想等 (樽井@mfeed編)

- 1,「経過時間」という切り口
  - 顧客対応を考えると、非常にリーズナブル。
  - グラフ上の「面積」で影響規模を測るというアプローチとも言える。
    - 一時的な spike であれば面積的にインパクトは小さいので無視させれば、早朝帯の誤検知 (FP) も抑制できるのではないかと思う。(ただ、面積で通知有無を判断する場合にもパラメータに悩む可能性がある。)
- 2, ベイジアンフィルタを応用
  - ツール(異常検知アルゴリズム)に正解を教え、育てていくアプローチが新鮮。
  - ある程度、自動的にフィルタが生成、かつ追従されて精度が保てることが条件ではあるが、ベイズの定理により生成されたルールをオペレータが fine tune (修正および無効化)できれば、事業者者毎の運用ポリシーも実装できそうな予感あり。
    - 例えば、同じような波形変化でも、リンクAは無視して良いが、リンクBは夜間でもエスカレーションして欲しい等のカスタマイズが可能に。
    - 「なるべく自動で‥、」とは考えているが、精度を考慮すると、最後はやはりオペレータの皮膚感覚に頼らざるを得ないだろうと思っている。
- 3, ツールの特性を知る、ということ。
  - ネットワークも冗長化構成になっているように、監視ツールも複数システムを動かす必要があるのではないか。その際、ツール毎の検知特性、強み、弱みという特徴を予め把握しておくことが見逃し(FN)は撲滅するためには重要になる。
    - ツールAでは検知しきれない部分を別のツールBでカバーできば見逃しが抑制できる。一方、誤検知の発生頻度も上がってしまう為、そこは引き続き悩ましくもある。

得られた知見、感想等 (原田@NTT-lab編)

- 1, ツールを評価することの難しさ.
  - 精度:「それって本当に通報すべき？」
    - １０人いれば１０通りのポリシーがありそう. 目指すべき異常検知の正解をどうやって決めれば良いか, 悩ましいところ.
    - １０人でも１００人でも, 全員を納得させられる通報をするアルゴリズムを作ることは理想だけど, そんなことはできるのだろうか. それよりも, オペレータの嗜好に合わせてカスタマイズできるアルゴリズムのほうが実現性がある？
  - 機能:「カレンダーを利用する？」
    - 波形を見て平日か休日かを判定する機能を閾値アルゴリズムに導入するよりも, カレンダーを取り込んで平日と休日を設定する別のツールを用意すべき？
    - 自力でどこまでできるかを試すのは面白いけど, ツールとしての完成は遠くなりそう.
- 2, 別分野の類似技術.
  - 高速フーリエ変換
    - 周期を見つけることは得意そう. 時間ができたら波形を入力して試してみたい. 一週間の周期を見つけるためには何週間分のトラヒックが必要なんだろうか.
    - 周期を見つけるのは閾値設定アルゴリズムの機能と言うよりも, 事前処理という位置づけかも. 閾値設定装置‥と考えるべき？
  - 電力需要
    - 需要曲線はイベントに大きく影響を受けるらしい. ネットワークトラヒックでも似たようなことは言えるかもしれない. 例えばルーティングテーブルが書き換わるとか？
    - いずれにせよ, トラヒックのボリューム変動だけを見ていては予測できないことも, いろいろな情報を使えば予測できるようになるかも知れない.

得られた知見、感想等 (谷津@KDDI編)

[JANOG発表前]

- ツールの評価方法が分からない。
  - 全標本中、正解標本の全てを抽出できているかどうかが分からない。
    - False Negative についての評価ができていないということ。
  - 画一な評価ができていないので、細かな変更の前後で精度が向上したのか後退したのか分からない。
    - code の変更ができない。

[JANOG発表後]

- 1, 周波数成分おもしろそう。
  - 1/300[Hz]でサンプリングしてるので、1/600[Hz]以上の高周波成分はノイズか。うっ。
  - 成分分析したいなら、周波数成分よりもflow成分の方が。
    - 成分分析などの時系列パターン以外の解析は今回の out of scope だったような。
- 2, ベイズの定理はおもしろそう。
  - 評価すべきコンテンツは、bps値(と時刻？)だけ。これだけでベイズの定理が適用できるのか？
- 3, 微小地震検出のアルゴリズムおもしろそう。
  - ふだんおとなしいパターンに適用するのに向いていそう。
  - 時間変化のないパターンに適用するのに向いていそう。
  - 決まった時間変化をするパターン、への応用を考えてみたらどうかな。

得られた知見、感想等 (廣川@NTT-lab編)

- 1, pps, fpsの監視を行ってみてはどうか。
  - bpsだけの監視では異常を検出するのは難しい。見つけたい異常と誤検出の区別がつかない。
  - pps, fpsの方がDoS攻撃による異常の特徴などが現れやすい。
- 2, 異常の継続時間によって、その異常の重要度を表してはどうか。
  - TLDは急激な変動が発生した箇所のみを通知するため、異常が徐々に収まると異常の終了を検出できない。そのため、異常の継続時間がわからない。
  - 異常終了通知の必要性は感じているので、検討したい。
  - 終了判断アルゴリズムは異常検知アルゴリズムとは切り離して考えて、どのアルゴリズムとも組み合わせられるようにしたい。
- 3, 休日、平日といったパターン分け自体を学習により自動生成する。
  - カレンダーを用意しなくても休日･祝日と平日を区別し異常検出を行える。
  - カレンダー通りではないトラヒックにも対応できる。
  - 実現は難しそうだが取り組んでみたい。
  - 処理が重くなりそうなので、複数波形に同時に適用するのは難しいかも。
- 4, トラヒック監視の需要について
  - 発表資料内データAのような特徴が現れるトラヒックには異常検知についての強い需要がある。
    - → 周期的に現れる急変動への対処は早めに行いたい。
  - 監視だけではなく、その後の原因調査も自動で行って欲しい、という需要もある。

謝辞

- JANOG21で発表することができて良かったです。
- 今後も活動を継続していきたいと考えています。
- 関係各位、ありがとうございました。

以上です

2008年2月
JANOG21 / Program Producer 樽井行保