5 統計的異常検出3手法

山西 健司（NEC インターネットシステム研究所）
k-yamanishi@cw.jp.nec.com

竹内 純一（NEC インターネットシステム研究所）
tak@ap.jp.nec.com

丸山 祐子（NEC インターネットシステム研究所）
y-maruyama@bp.jp.nec.com

なぜ異常検出は重要か？

大量なデータを手に入れると，通常そこには多かれ少なかれパタンが見えてくる．そのようなパタンから大きく外れた異常なデータを発見すること，あるいはパタンの異常な変化を検出することを，「異常検出」(Anomaly Detection) と呼ぶ．これはデータマイニングの重要な機能の1つである．

たとえば，クレジットカードのトランザクションからの異常検出問題を考える．Aさんのクレジットカードのトランザクションには，Aさんがいつ，どこで，いくらの買い物をしたのかが記されている．大量のトランザクションからはAさんの利用パタンが見えてくるだろう．しかし，誰かがAさんになりすまして利用したとすれば，Aさんの利用パタンとは異なるトランザクションが出てくる，あるいはその利用パタン自身が急に変化してしまうかもしれない．そのような異常を検出することは，Aさんのクレジットカードの**不正検出 (Fraud Detection)** を行うための1つの手段を与えるのである．

ほかにもたとえば，以下のような応用分野が考えられる．

- **ネットワーク/コンピュータセキュリティ：**
  ネットワークのアクセスログやIDS (Intrusion Detection System) のログからの異常検出によって，新種のウイルスやワームの発生を検出する．あるいはUNIXのコマンド履歴からの異常検出により，なりすましや情報漏えい等の内部犯罪を検出する．
- **ネットワーク/コンピュータの障害検出：**
  SYSLOGとよばれる装置のメッセージデータ等からの異常検出により，ネットワークの障害やその予兆を検出する．

以上のように，異常検出の応用分野は多岐にわたり，いずれもIT社会の重要な問題に結びついている．

統計的な異常検出

異常検出の代表的な方法として「**統計的な異常検出**」が挙げられる．これはデータの生成機構が統計的モデル（確率分布）で表現できると仮定した場合の異常検出の方法論である．本稿ではこの方法論をベースとして統一的な異常検出の枠組みについて解説する．まず，統計的な異常検出は，基本的には以下のステップにまとめることができる．

ステップ1： これまでに得られたデータから，データ発生分布の統計的モデルを学習する．

ステップ2： ステップ1で学習されたモデルを基に，異

機能	統計的モデル	検出対象	応用	文献
外れ値検出	独立モデル（ガウス混合分布，ヒストグラム）	モデルに相対的な外れ値	不正検出 進入検出	1) 5)
変化点検出	時系列モデル(ARモデル，回帰モデル)	時系列上の急激な変化	ウイルス／ワーム検出	2) 4)
異常行動検出	行動モデル（混合隠れマルコフモデル，ベイジアンネット）	異常なセッション／行動パタン	なりすまし検出，障害検出	3) 6)

表-1　統計的異常検出3手法

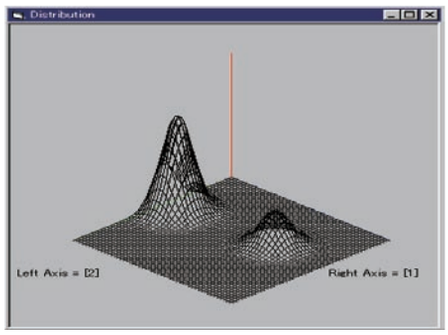

図-1　ガウス混合分布の確率密度関数

常なデータ，またはモデルの異常な変化を検出する．

統計的な異常検出では，ステップ1でどのような統計的モデルのクラスを対象にするか？　ステップ2でどのような異常を検出するか？　といった観点からさまざまなバリエーションが生まれる．本稿では，**表-1**のように異常検出を「外れ値検出」，「変化点検出」，「異常行動検出」の3手法に分類して解説することを試みる．これは順に，独立モデル→時系列モデル→行動モデルと，より「ダイナミックな」異常を検出する問題へと進んでいく，異常検出の1つの切り口である．それぞれを表-1にまとめる．

データマイニングとして統計的異常検出手法を設計する際，高い検出精度を達成することはもちろん重要であるが，これに加えて以下の要件を満たす必要がある．

(1) (効率性) オンライン形式で実時間で検出を行う
(2) (適応性) データ源の性質が時間とともに変化する非定常な状況に対しても適応的である

従来の統計的手法では，これらは必ずしも考慮されなかった．本稿では，これらを満たす異常検出について述べる．

適応的な外れ値検出

外れ値検出とは，他のデータ群から著しく離れた異常値をとるデータを検出することである．今，仮にデータは1次元であるとすると，外れ値検出の最も典型的な方法は，これまで得られたデータの平均値をμ，標準偏差をσとして，新しい入力データxがμより3σ以上離れていれば，xを外れ値と見なして検出する**閾値判定法**である．つまり，xが外れ値である判定基準は以下で与えられる．

$$|x - \mu| > 3\sigma$$

しかし，そのような方法を適用する背景には，「データの発生分布が定常的に同一のガウス分布$N(\mu, \sigma)$で与えられる」という仮定が置かれている．

ところが，実際のデータマイニングの状況では，(1) データの発生分布は時間とともに変化し(非定常性)，(2) データの発生分布は単純な正規分布で表されるものではない．しかも，(3) データは多次元で，(4) データが与えられるごとにオンラインで外れ値検出を行わなければならない．そのような状況に対応して，一般的な適応的外れ値検出を行うためのエンジンSmartSifterが，Yamanishi et al.[5]によって提案されている．これを以下で紹介する．

SmartSifterは，オンラインで統計的モデルを適応的に学習し，そのモデルに対する各データの異常度合いをスコアリングするというステップを踏む．以下詳細を与える．

統計的モデル：データ発生の統計的モデルとして独立分布を仮定する．多次元連続値データを仮定し，ガウス混合モデルと呼ばれる有限個のガウス分布の線形の重ね合わせで表現する(**図-1**)．これはxを確率変数，kを与えられた正整数として，以下のような確率密度関数で表現される．

$$p(x \mid \theta) = \Sigma_{i=1}^{k} c_i \, p(x \mid \mu_i, \Sigma_i).$$

$p(x \mid \mu_i, \Sigma_i)$は$i$番目の混合成分である平均$\mu_i$，分散行列$\Sigma_i$のガウス分布を表す．$c_i$はその重みを表し，$c_i > 0$かつ$\Sigma_{i=1}^{k} c_i = 1$を満たすとする．すべてのパラメータを

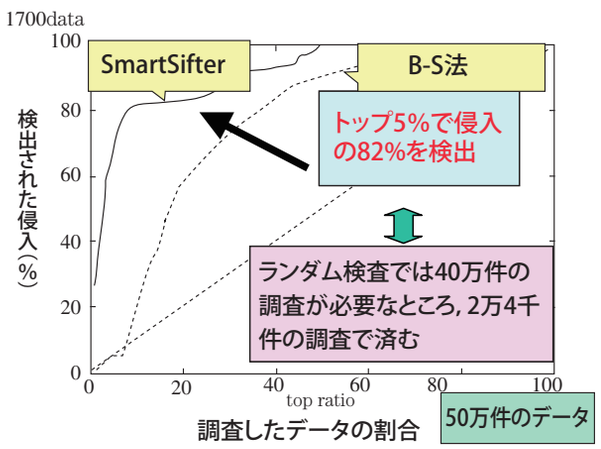

図-2　SmartSifterによる侵入検出

まとめて，ベクトルθで表現する．

$$\theta = (c_1, \mu_1, \Sigma_1, \ldots, c_k, \mu_k, \Sigma_k).$$

ガウス混合分布は，データがどの混合成分から発生したかという実際には観測されない変数zを導入することにより(これを隠れ変数と呼ぶ)，観測される変数xと隠れ変数zの同時分布として以下の形で表現することもできる．

$$p(x, z = i \mid \theta) = c_i p(x \mid \mu_i, \Sigma_i)$$

学習：これらの統計的モデルのパラメータθの値を，データを逐次的に取り込むごとに**オンライン忘却型学習アルゴリズム**を用いて学習する．これは過去のデータほどその影響を徐々に少なくしながら統計的モデルを学習することで，適応的な学習を実現したアルゴリズムである．

以下，隠れ変数を伴うモデルに対して，オンライン忘却型アルゴリズムを一般的な形式で与えよう．

今，xを観測される確率変数，zを隠れ変数として，k次元パラメータベクトル$\theta = (\theta_1, \ldots, \theta_k)$で指定される統計的モデル$p(x, z : \theta)$を学習したいとする．ここで

$$[x\text{の周辺分布}] \quad \int p(x, z : \theta)\, dz = p(x : \theta),$$
$$[z\text{の事後確率}] \quad p(z \mid x : \theta) = p(x, z : \theta) \,/\, p(x : \theta)$$

であることに注意する．以下のステップにより，各時点tでθの推定値$\hat{\theta}^{(t)}$を計算する．

[初期化]　$t := 1, \hat{\theta}^{(0)}, F^{(0)}(\theta)$ (k次元ベクトル): given
$0 < r < 1$　忘却係数
($t := 2, \ldots, n$ 以下のE-ステップとM-ステップを繰り返す)
[E-Step] x_tを読み込んで以下の関数を定義する．

$$F^{(t)}(\theta) := (1-r)\, F^{(t-1)}(\theta) + r \int p(z \mid x_t : \hat{\theta}^{(t-1)}) \log p(x_t, z : \theta)\, dz$$

[M-Step] $\hat{\theta}^{(t)} := \arg\max_\theta F^{(t)}(\theta)$

ここでのポイントは，$\hat{\theta}^{(t)}$は時点jの重み付き十分統計量である$\int p(z \mid x_j : \hat{\theta}^{(j-1)}) \log p(x_j, z : \theta)\, dz$を$j$に関して，$r(1-r)^{t-j}$の重みを与えて平均し，過去のものほど指数的に効果を小さくするような対数尤度関数の極大値として求められるということである．したがって，rを1に近づけるほど忘却の効果は大きくなる．

ガウス混合分布に特化したオンライン忘却型学習アルゴリズムの具体的形式は文献5)を参考にされたいが，その場合，データ長nに対して$O(n\, k^2)$の計算量を要する．

スコアリング：各データx_tのスコアは次式のようなそれまで学習されたモデル$p(x \mid \hat{\theta}^{(t-1)})$に対するShannon情報量で計算する．

$$-\log p(x_t \mid \hat{\theta}^{(t-1)}).$$

よって，スコアが高いほど異常値度合いが高いと見なせる．

■外れ値検出の不正侵入検出への応用

KDDCup99 (http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html) と呼ばれるデータマイニングのコンテストに用いられ，一般に公開されているネットワークログのベンチマークデータを用いてSmartSifterによるネットワーク不正侵入検出の実験を行った[6]．データの属性としては(duration, src_bytes, dst_bytes)の3つを用いた(durationは接続時間，src_byteはサーバへ送った情報量，dst_byteはサーバから受信した情報量を表す)．実験にはlog inに成功した50万件のデータを用いた．そのうち侵入に成功したデータは0.35%含まれていた．

図-2は，全データをスコア順にソートしたときにスコア上位のデータ中にどれだけの侵入が実際に含まれていたかを示している．横軸はスコアの高い順に取り出したデータの割合を表し，縦軸は全侵入に対する検出された侵入の割合(カバー率)を表す．実線はSmartSifterの性能を，破線は競合方式であるBurge & Shawe-Taylorの自己組織化マップに基づく方法[1]の性能を表している．

このグラフからSmartSifterの性能が競合方式のそ

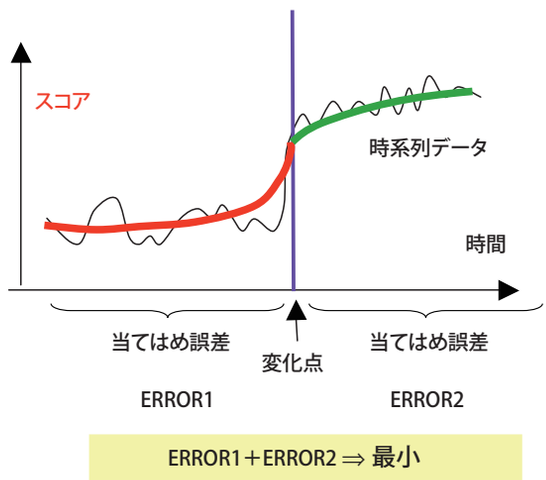

図-3　統計的検定に基づく変化点検出

れを大きく上回っていることが分かる．また，全侵入の8割を検出するために，ランダムに抽出した場合は全体の8割のデータを調べる必要があるのに対し，SmartSifterによるスコアで優先順位を付けて調べた場合には全体の5%のデータを調べれば済むことが分かる．このように外れ値検出は不正検出のための調査工数を激減させる(この例の場合は10分の1以下に削減する)効果を持つ．

統計的検定に基づく変化点検出

前出の外れ値検出では，データの発生モデルは独立分布であると仮定していた．しかしながら，実際には，データが時系列をなす場合があり，時系列的な性質が急に変化した時点(変化点)を検出しなければならないことがある．そのような問題を**変化点検出**と呼ぶ．

たとえば，ネットワークのFirewallでのアクセスドロップ数(アクセスが拒否されたログの数)の時系列データを観測する場合，ウイルスやワームの発生によりある時点を境目に集中的にその数が増加する傾向にある．そのような時系列の変化点を検出することにより，ウイルスやワームの発生を早期に検出することが期待できる．

変化点検出の最も基本的な方式は，Guralnik and Srivastava[2]の仕事に見られるように，データ列にARモデルや多項式回帰モデルなどの時系列モデルを当てはめていき，変化点の候補となる点の前後で別々に時系列モデルを当てはめた場合が，そうしない場合に比べて当てはめ誤差を有意に少なくできるかどうかを検定し，YESならば，その候補点を変化点と見なすという方式である(図-3)．

これを定式化しよう．データ系列$x_1^n = x_1,..., x_n$の変化点候補をtとして，tの前後の時系列データをそれぞれ$x_1^t = x_1,..., x_t$, $x_{t+1}^n = x_{t+1},..., x_n$と書く．ここに各$x_j$は多次元連続値をとるとする．時系列モデルとしては，たとえば，k次の自己回帰モデル(AR(Auto Regression)モデル)を用いる．これは$x_{t-k}^{t-1} = x_{t-k}, ..., x_{t-1}$が与えられたときの$x_t$の条件付確率密度関数が以下で与えられるモデルである．

$$p(x_t \mid x_{t-k}^{t-1}) = \frac{1}{\sqrt{2\pi \mid \Sigma \mid}} \exp\left(-\frac{(x_t - w_t)^T \Sigma^{-1} (x_t - w_t)}{2}\right)$$

ここに

$$w_t = \Sigma_{i=1}^{k} \alpha_i (x_{t-i} - \mu) + \mu$$

であり，$\theta = (\alpha_1,..,\alpha_k, \mu, \Sigma)$は実数値パラメータである．各パラメータはデータから最尤法などにより推定するとし，推定されたパラメータを用いて上式によって計算されるw_tの予測値を$\hat{w}_t$と書くとき，モデルのデータx_1^nに対する当てはめ誤差$I(x_1^n)$を二乗誤差関数で計算する．

$$I(x_1^n) = \Sigma_{t=1}^{n} (x_t - \hat{w}_t)^2$$

そこで，$I(x_1^t) + I(x_{t+1}^n)$のtに関する最小値$t = t^*$が，δを与えられた閾値として，判定基準

$$(I(x_1^n) - (I(x_1^{t^*}) + I(x_{t^*+1}^n))) / (n - t^* + 1) > \delta$$

を満たすならば，$t = t^*$は変化点であると見なす．

変化点が見つかればその前後で同様のプロセスを再帰的に繰り返していく．

2段階学習に基づく変化点検出

前章で紹介した方式は，一般に計算時間がかかり，大量のデータを扱うデータマイニングとしては重すぎるアルゴリズムになっている．そこで，よりシンプルでオンライン処理に向いた変化点検出アルゴリズムを実装したエンジンが，Yamanishi and Takeuchi[4]により提案されている．これは，時系列データを2段階に渡って学習し，各時点の変化点スコアをオンラインで算出するものである(図-4)．このエンジン(以下ChangeFinderと書く)の

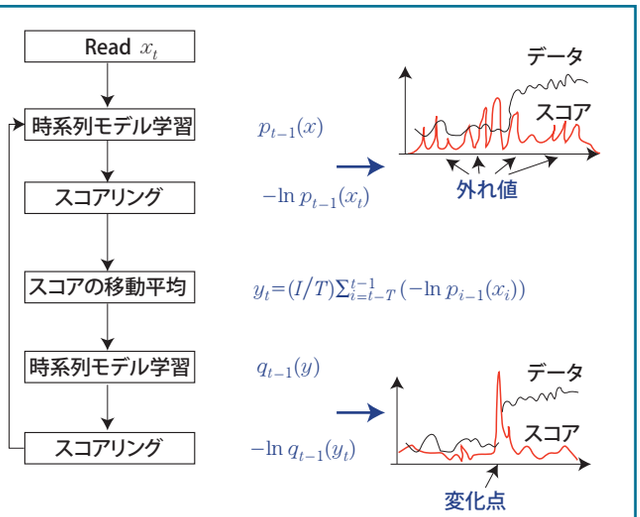

図-4　2段階学習に基づく変化点検出

図-5　ChangeFinderを用いたワーム発生検出

原理を以下に示そう．

第1段階学習：前述のARモデルのような時系列モデルを用いて，各時刻tに対して，これを前述のオンライン忘却型学習アルゴリズムと同様の方式で学習する．学習した確率密度関数を$p_t(x)$とする(注意：ARモデルは定常性を仮定する統計的モデルであるが，忘却型学習によって非定常性も扱うことができる．具体的アルゴリズムは文献4)を参考にされたい)．各時刻tについてデータx_tの外れ値スコアをShannon情報量$-\log p_{t-1}(x_t)$として計算する．

平滑化：一定サイズTのウインドウ内のデータの外れ値スコアの平均$y_t=(1/T)\sum_{i=t-T}^{t-1}-\ln p_{i-1}(x_i)$を計算し，ウインドウをスライドすることによって移動平均スコアの時系列$\{y_t : t=1, 2, ...\}$を構成する．

第2段階学習：この時系列に対して再度ARモデルのような統計的モデルを当てはめ，これを学習して，その系列を$q_t(y)$とする．各時点tにおいてShannon情報量$-\ln q_{t-1}(y_t)$を時刻tの変化点スコアとして計算する．変化点スコアが大きいほど，tが変化点である度合いが高い．

ChangeFinderの鍵は，第1段階学習では時系列中の外れ値しか検出できないところを，外れ値スコアの平滑化を通じて，本質的なモデルの変動を検出しているところにある．

計算量としては，データ数nに関しては，統計的検定に基づく方式が$O(n^2)$であるのに対して，ChangeFinderの計算量は$O(n)$で済むので，後者の方がはるかに効率的である．また，ChangeFinderでは，データの発生分布の分散が突然変化する場合でもこれを検出できるが，従来手法では必ずしもこれを検出できないことが確認されている．

■変化点検出のワーム検出への応用

変化点検出はネットワークのアクセスデータからのワーム検出に応用できる．**図-5**にChangeFinderを用いてMS.Blastの発生を検出した結果を示す．図-5は，脆弱性が発見されたポート135への単位時間当たりのアクセスドロップ数(発生頻度)の時間的変化と，これを入力としてChangeFinderが計算した変化点スコアの時間的曲線を示している．ChangeFinderのスコアには2カ所の鋭いピークが現れているが，これらはMS.Blastの1次発生と2次発生の初期に対応している．2段階に渡って発生したといわれるMS.Blastの発生の初期段階を捉えていることが分かる．このように，変化点検出は未知ワームや攻撃などの出現を早期に検出するのに有効である．

■異常行動検出

一連の行動履歴を示す時系列データを**セッション**と呼ぶ．大量のセッションデータを入力として，異常なセッション，または異常な行動パタンの出現を検出する問題を**異常行動検出**と呼ぶ．たとえば，ある人の一定時間内のUNIXのコマンド履歴系列をセッションと見なすと，大量のセッションデータから異常なセッションを発見す

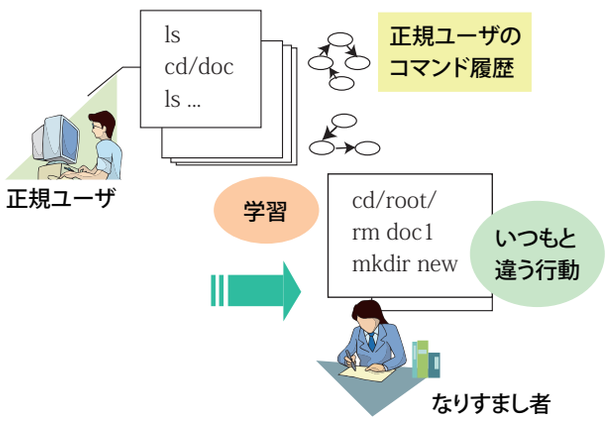

図-6　異常行動検出によるなりすまし検出

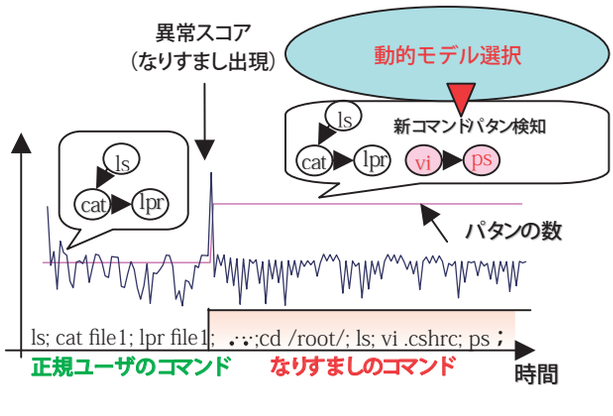

図-7　動的モデル選択によるなりすましパタン検出

ることにより，なりすまし犯罪行為を検出することができる．これは一種の異常行動検出問題である(**図-6**)．変化点検出では時系列データの局所的な異常を検出するのに対して，異常行動検出では大局的な行動パタンとしての異常を検出することを目的としている．

松永，山西[6]は情報理論的手法に基づいて異常行動検出のためのエンジン(以下，AccessTracerと呼ぶ)を提案している．以下に文献6)に従って，そのアルゴリズムの概要を示そう．

統計的モデル：セッションのデータ列の発生分布を隠れマルコフモデル(Hidden Markov Model (HMM))の有限混合分布を用いて表現する．隠れマルコフモデルは，確率的な状態遷移を表現するのに適した確率モデルである．ここでは，1つの行動パタンを1つのHMMで表し，複数の行動パタンが表れる可能性があることを，それらの線形結合，すなわち複数のHMMの混合分布によって表す．

たとえば，コマンドなどの有限アルファベット上に値を持つ長さNのセッションを$x^N=x_1,\ldots,x_N$と書くとき，x^Nの生成モデルは以下で表される．

$$P(x^N)=\sum_{i=1}^{k} c_i P_i(x^N).$$

c_iはi番目の混合成分の重みを表し，$c_i>0$かつ$\sum_{i=1}^{k} c_i=1$を満たすとする．$P_i(x^N)$は隠れマルコフモデルを表す．$P_i(x^N)$はZを隠れ変数(状態数をMとする)，$a(Z\mid Z')$を隠れ変数の1次の状態遷移行列，$b(X\mid Z)$を状態から出力への遷移行列，$\gamma(Z)$は状態Zの初期確率とすると以下のように書ける．

$$P_i(x^N)=\Sigma\gamma(z_1)\,\Pi_{j=2}^{N} a(z_j\mid z_{j-1})\,\Pi_{j=1}^{N} b(x_j\mid z_j).$$

ただし和はすべての長さNの状態列$(z_1,\ldots,z_N)$の組合せについてとられるものとする．

学習：HMMの混合分布をSmartSifterと同様なオンライン忘却型学習アルゴリズムで学習する．忘却効果により行動パタンが変化しても適応して学習することができる．

スコアリング：学習されたモデルに基づいて各セッションの異常スコアを計算する．異常スコアは，外れ値検出や変化点検出と同様にShannon情報量によって計算するが，セッションの長さで正規化するものとする．

動的モデル選択による異常行動パタン検出

HMMの有限混合分布の最適な混合数は本質的に異なる典型的な行動パタンの数を示している．この最適な混合数をデータが与えられるごとに逐次的に求めていくことを「**動的モデル選択**」と呼ぶ[3), 6)]．最適な混合数の変化は統計的モデルの構造的な変化を意味する．動的モデル選択は，これをトラッキングすることにより異常検出を行おうとする試みである．

UNIXのコマンドラインの解析例をとりあげよう．10コマンドを1つのセッションとしてセッションの時系列を構成した．**図-7**では横軸がセッションの番号を表し，小刻みの折れ線は各セッションの異常スコアのグラフを表し，階段状の折れ線はHMMの混合数の最適値を動的モデル選択で求めた値の変化グラフを表している．なりすましなどの異常行動の出現に伴い，コマンド履歴に新しい行動パタンが生成され，混合数が増えているのが検出されている．このように動的モデル選択は，新しいパタンの生成や消滅を検出する上で有効である．

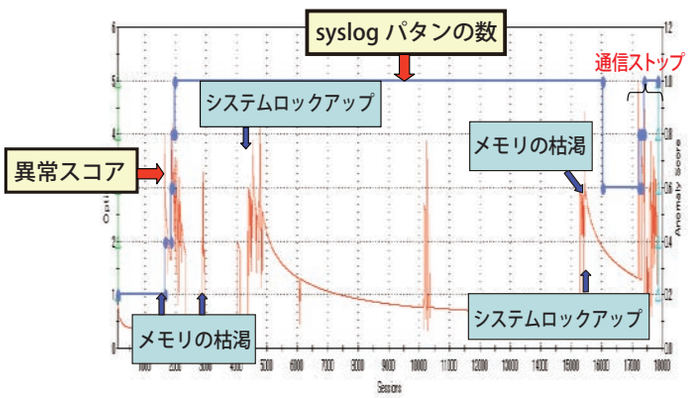

図-8　AccessTracerによる障害検出

文献6）に従って動的モデル選択の数学的原理を簡単に示す．k個の混合数を持つHMMの有限混合モデルについて，時刻$j-1$までの全セッションから学習されたモデルを$p^{(j-1)}$と書くとき，セッション列$\boldsymbol{x}^t = \boldsymbol{x}_1,.., \boldsymbol{x}_t$ $(\boldsymbol{x}_j = x^{N_j} = x_1,.., x_{N_j})$に関する混合数$k$のモデルに関する**予測的確率的コンプレキシティ**を次式で定義する．

$$I_k(\boldsymbol{x}^t) = \Sigma_{j=1}^{t} - \log p^{(j-1)}(\boldsymbol{x}_j).$$

これは情報理論的には$\boldsymbol{x}^t = \boldsymbol{x}_1,.., \boldsymbol{x}_t$をデータが与えられるごとに逐次的に歪無し符号化する際の総符号長を意味している．動的モデル選択は**記述長最小原理**と呼ばれる情報理論の原理に基づいて，予測的確率的コンプレキシティを最小になるような混合数kの値を各時刻tにおいてセッション$\boldsymbol{x}_t$を入力するごとに逐次的に選択していくアルゴリズムである．さらに，混合数の間に遷移確率を導入して混合数の最適な系列をよりダイナミックに検出するアルゴリズムも提案されている[3]．

■異常行動検出の障害検出への応用

AccessTracerは，コンピュータのSyslogからの障害検出にも応用することができる．**図-8**はネットワークを構成する装置のSyslogセッションの時系列に対して出力した異常スコアのグラフである．システムのロックアップが2回起った後に通信ストップが起っているが，これらに対応して高いスコアが出ている．システムロックアップの直前にはその予兆としてメモリ枯渇があったが，これらに対しても高いスコアが与えられていた．このように，異常行動検出はシステムの異常な振る舞いを早期に検出するのに有効である．

■異常検出の今後

本稿で紹介した異常検出技術の方法論は，統計的モデルのさまざまなバリエーションに展開することによって，セキュリティや障害検出分野のみならず，金融や医療分野におけるリスク管理やなど広い範囲の異常検出問題に適用することができると期待できる．

本稿では，異常なデータを検出するばかりでなく，データの背後にある統計的モデルの構造的な変化を検出する考え方（動的モデル選択）をも示した．後者は，異常検出の新しい考え方であるとともに今後ますます重要になると思われる．また，それは将来的には異常の「予兆」を検出する方法論と結びついていくであろうと夢は膨らむのである．

参考文献

1）Burge, P. and Shawe-Taylor, J.: Detecting Cellular Fraud using Adaptive Prototypes, in Proceedings of AI Approaches to Fraud Detection and Risk Management, pp.9-13 (1997).
2）Guralnik, V. and Srivastava, J.: Event Detection from Time Series Data, in Proceedings of the Sixth ACM SIGKDD International Conference on Data Mining and Knowledge Discovery, ACM Press, pp.32-42 (1999).
3）Maruyama, Y. and Yamanishi, K.: Dynamic Model Selection and Its Applications to Computer Security, in Proceedings of the IEEE Information Theory Workshop 2004, http://ee-wcl.tamu.edu/itw2004/program.html
4）Yamanishi, K. and Takeuchi, J.: A Unifying Approach to Detecting Outliers and Change-Points from Non stationary Data, in Proceedings of the Eighth ACM SIGKDD International Conference on Data Mining and Knowledge Discovery, ACM Press, pp.676-681 (2002).
5）Yamanishi, K., Takeuchi, J., Williams, G. and Milne, P.: On-line Unsupervised Oultlier Detection Using Finite Mixtures with Discounting Learning Algorithms, Data Mining and Knowledge Discovery, Kluwer Academic Press, 8, pp.275-300 (2004).
6）松永，山西：情報理論的手法に基づく異常行動検出，第2回情報科学技術フォーラム予稿集，pp.123-124 (2003).

（平成16年12月3日受付）