L-002

主成分分析を用いた分類器によるSQLインジェクション攻撃の自動検出法

Automatic Detection Method of SQL Injection Attack by a Classifier using Principal Component Analysis

園田道夫 † Michio Sonoda　松田健 ‡ Takeshi Matsuda　小泉大城 § Daiki Koizumi　趙晋輝 † Chao Jinhui

1. はじめに

近年，スマートフォンやタブレット PC などの普及に伴い，Web の仕組みを軸としたシステムの社会的な需要が急速に増している．それと同時に，Web の入力フォームからデータベースを直接操作する SQL インジェクション攻撃 [1] を初めとする Web アプリケーションの脅威も増大しつつあり，不正アクセス事件が後を絶たないことが社会問題となっている．SQL インジェクション攻撃については，システムの開発時に手順に則ってプログラムを作成するだけでこ技術対策が確立されているが，そのような技術的対策を開発時に取り入れられていないシステムには未だ致命的な脆弱性が存在する状況が続いている．こうした攻撃からシステムを守るために，過去の攻撃パターンをブラックリスト化し Web のフォームから入力された文字列パターンとリストを照合する Web アプリケーションファイアウォールが導入されている．しかしながら，Web アプリケーションファイアウォールには

1. リストの照合による攻撃検出方法では新種の攻撃の検出が出来ないこと
2. 新種の攻撃が出現した時、検出可能なリストを作成するまでの間にその攻撃に対応することが出来ないこと
3. 新種の攻撃が絶えず開発され続けていること

などの問題点が挙げられる．このような問題に対し，Support Vector Machine (SVM)[2] などの機械学習の手法を用いて SQL インジェクション攻撃を検出する手法が研究されており [3][4][5]，著者らも，SQL インジェクション攻撃の文字列に頻出するセミコロンやシングルクォートなどの記号を攻撃特徴文字として SQL インジェクション攻撃を検出するアルゴリズムを提案している [11][12]．SVM は汎化能力が高い機械学習として広く利用されているが，データの特徴空間の定義の仕方によっては汎化能力にも影響が現れるものと考えられる．そこで本研究では，主成分分析を用いた分類器を構成し SQL インジェクション攻撃を検出する手法を提案し，SVM と提案手法の汎化能力に関する考察を行う．主成分分析はデータの次元縮退のために利用されており，本研究においても SQL インジェクション攻撃の攻撃特徴を抽出する目的でデータの特徴空間に対して主成分分析を適用した．

† 中央大学大学院理工学研究科
‡ 静岡理工科大学総合情報学部コンピュータシステム学科
§ サイバー大学 IT 総合学部

本章に続く第 2 章では，SVM の概要とそれを応用した従来研究についてまとめ，第 3 章では，2 種類のアルゴリズムを提案する．次なる第 4 章では，2 種類の提案アルゴリズムを人工データによって性能評価を行った結果について述べる．最後の第 5 章では評価結果の考察とまとめを行う．

2. 従来研究

本章では Support Vector Machine (SVM) [2] について簡単に紹介し，SVM を応用している従来研究についてまとめを行う．

2.1.SVM の概要

$\mathbf{R}^n$ 上にある 2 つの異なるグループ C_1, C_2 に分類可能な m 個の点 $\mathbf{x}_i \in \mathbf{R}^n$ $(i = 1, 2, \cdots, m)$ が与えられたとする．もし $\mathbf{x}_i \in C_1$ なら $y_i = 1$，$\mathbf{x}_i \in C_2$ なら $y_i = -1$ とすると，m 個の点の集合は

$$D_m = \{(\mathbf{x}_i, y_i) \in \mathbf{R}^n \times \{1, -1\} | i = 1, 2, \cdots, m\}$$

と表すことができ，D_m をサンプルと呼ぶこととする．もし D_m を構成する点達が $\mathbf{R}^n$ の超平面 $H = \{\sum_{j=1}^{n} a_j t_j + b = 0\}$ で完全に 2 つのグループ C_1, C_2 に分けることができるとき，サンプル D_m は線形分離可能であるといい，H を分離平面という．ここで $\{t_1, t_2, \cdots, t_n\}$ は $\mathbf{R}^n$ の座標を表すものとする．また，線形分離可能でない場合は，非線形な関数であるカーネル関数を用いてサンプルを線形分離可能にできる場合も存在する．しかし，サンプルが線形分離可能であっても，一般的に分離平面の決め方は一意に定まるものではない．SVM はサンプル D_m から分離平面 H を求める手法であるが，H とサンプル D_m の $x_i \in \mathbf{R}^n$ の点 $\mathbf{x}_i$ との距離を最大にする $a_1, a_2, \cdots, a_n$ を求めることで汎化能力を高くすることができるという特徴をもっている．また，$\mathbf{R}^n$ のすべての点 $\mathbf{x}_i$ が $a_1, a_2, \cdots, a_n$ の決定に関わるわけでなく，$a_1, a_2, \cdots, a_n$ の決定に関係のある点のことをサポートベクターといい，サポートベクターの点の情報から分離平面 H の b の値を求めることができる．

2.2.SVM の応用研究

前節で述べたように，SVM はデータを 2 種類のパターンに分類できる問題に適用される機械学習の一方法で，その他いくつか存在する分類法の中でも分類の精度を高く実現できる手法として知られており，情報工学，医療情報，社会工学などの分野で広く利用されている．

その中でも情報セキュリティ，特に本論文で扱う SQL インジェクション攻撃に着目すると，この種の攻撃は Web アプリケーションを導入する際にサニタイズやプ

リペアドステートメントといった技術的対策を行うことでこの攻撃の被害に遭遇する可能性は低くなる．しかしながら，実際には SQL インジェクション攻撃の被害報告数は増加傾向にあり，Web アプリケーションを SQL インジェクション攻撃から防御する手段として過去に観測された実際の攻撃のパターンをブラックリスト化したり，パターン認識や SVM のような機械学習と呼ばれる手法を用いて攻撃を検出する方法 [3][4][5] などが提案されている．

2.3. 主成分分析の概要

主成分分析は，データの特性が多変量で表現されているとき，つまりデータが高次元のベクトルで表されているような場合に，データの重要な特性を損なわずにより低い次元のベクトルでそのデータを表現可能なように変換するための手法である．以下，主成分分析の手法について簡単にまとめておく．$\mathbf{x}_i \in \mathbf{R}^n\ (i = 1, 2, \cdots, m)$ の座標を $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{in})$ とおく．サンプル D_m から得られる n 次正方行列

$$V = \frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} x_{i1}-\overline{x}_1 \\ x_{i2}-\overline{x}_2 \\ \cdots \\ x_{in}-\overline{x}_n \end{pmatrix}\begin{pmatrix} x_{i1}-\overline{x}_1 \\ x_{i2}-\overline{x}_2 \\ \cdots \\ x_{in}-\overline{x}_n \end{pmatrix}^t$$

を計算する．ここで $\overline{x_j} = \frac{1}{m}\sum_{i=1}^{m} x_{ij}\ (j = 1, 2, \cdots, n)$ であり，行列 V は分散・共分散行列と呼ばれる対称行列である．対称行列の固有値はすべて実数であるから，V の固有値を大きい順に $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ と並べ変え，対応する固有ベクトルを $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ とする．さらに対称行列において，異なる固有値に対応する固有ベクトル同士の内積は 0 になることから，V の固有ベクトル $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ を列ベクトルとしてできる行列 $U = (\mathbf{u}_1\mathbf{u}_2\cdots\mathbf{u}_n)$ は直交行列となり，行列 V は行列 U を用いて

$$V' = U^tVU = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \ddots & \\ \cdots & \cdots & & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

と対角化することができる．ここで U^t は U の転置行列であり，サンプル D_m から得られる列ベクトル

$$\begin{pmatrix} x_{i1}-\overline{x_1} \\ x_{i2}-\overline{x_2} \\ \cdots \\ x_{in}-\overline{x_n} \end{pmatrix}$$

に U^t を左側から作用させると新しい変数 $\{w_1, w_2, \cdots, w_n\}$ が得られる．

$$\begin{pmatrix} w_{i1} \\ w_{i2} \\ \cdots \\ w_{in} \end{pmatrix} = U^t\begin{pmatrix} x_{i1}-\overline{x_1} \\ x_{i2}-\overline{x_2} \\ \cdots \\ x_{in}-\overline{x_n} \end{pmatrix}$$

したがって V の固有ベクトルからできる直交行列 U は座標軸の変換を行う行列であると考えられる．固有ベクトルの示す方向を主軸といい，対応する固有値の値を主値という．また，行列 V' は $\{w_{i1}, w_{i2}, \cdots, w_{in}\}$ が対角行列であることを表しており，このようにデータから互いに無相関な因子を取り出し，データを無相関な因子の一次結合で表現することを主成分分析といい，分散・共分散行列が対角行列になるように主軸方向に座標を変換することを主軸変換という．

2.4. 主成分分析の先行研究

主成分分析による攻撃検知の研究はいくつか存在する。前述の通り重要な特徴を失わずに低次元に変換する手法であるので、多様な複雑さを含むデータの解析に適用されている例が多い。ネットワークトラフィックの解析を行ってサービス使用不能（DoS）攻撃を検知する研究や [6]、TCP や IP というレベルでの量的変化をベースに解析を行う研究などがある [7]。こうした攻撃検知研究においては多くの場合、異常検知という考え方が用いられているが、TCP や IP レベルでの量的変化や、通信の外形的特徴をもとにした検知では、Web アプリケーションへの攻撃などの通信の量よりも中身そのものが問題になる攻撃の場合には適用が難しい。また、先行研究の多くが対象としている攻撃は、DARPA データセット [8] やコンテストで用いられた KDD Cup data 99 というデータセット [9]、その進化版である NSL-KDD データセット [10] などに定義されている。これらのデータセットの定義は 1998 年から 2000 年までで行われていて、Web アプリケーションへの攻撃が多数を占める現在の攻撃状況を全て反映しているとは言い難い。

3. 提案アルゴリズム

著者らは半角スペースやセミコロンなどの記号を用いて SQL インジェクション攻撃の特徴を抽出するアルゴリズム [11][12] を提案しており，SQL インジェクション攻撃に頻出する記号を攻撃特徴文字と呼んでいる．本章では，攻撃特徴文字を用いてデータを分類する分類器を構成する方法を提案し，次の章で攻撃特徴文字の出現頻度を特徴空間とした SVM で SQL インジェクション攻撃を検出した場合と提案方法で検出した場合について考察を行う．著者らの従来研究では，SQL 文法によく利用される 20 文字の記号の中から SQL インジェクション攻撃特徴となる，

- 半角スペース
- セミコロン
- ダブルクォーテーション ZHANG
- 右側丸括弧
- 左側丸括弧

の 5 文字を攻撃特徴文字として抽出した [11][12]．これらの攻撃特徴文字を表 1 に示す．

本研究では，この 5 文字の攻撃特徴文字の入力文字列における出現頻度 x と残りの 15 文字の出現頻度 y を SQL インジェクション攻撃の特徴空間とし，データを分類する方法を以下のように提案する．提案アルゴリ

表 1: SQL インジェクション攻撃の攻撃特徴文字

番号	文字
1	半角スペース
2	セミコロン (;)
3	シングルクォーテーション (′)
4	右側丸括弧 ())
5	左側丸括弧 (()
6	右側中括弧 (})
7	左側中括弧 ({)
8	右側大括弧 (])
9	左側大括弧 ([)
10	シャープ (#)
11	パーセント (%)
12	ダブルクォーテーション (″)
13	アンバサンド (&)
14	バックスラッシュ (\)
15	パイプ (|)
16	等号 (=)
17	大なり不等号 (>)
18	小なり不等号 (<)
19	アスタリスク (∗)
20	スラッシュ (/)

ズムは以下の 2 つであり，アルゴリズム 1 では特徴空間に主成分分析を適用して分類器を構成し，アルゴリズム 2 では特徴空間に分布する 2 つのグループの平均の座標を用いて，データを縮約してから分類器を構成するものとなっている．

3.1. 提案アルゴリズム 1

(1) 学習用データの準備

- 攻撃データ n 個
  $\{(x_{A1}, y_{A1}), (x_{A2}, y_{A2}), \cdots, (x_{An}, y_{An})\}$
- 正常データ m 個
  $\{(x_{N1}, y_{N1}), (x_{N2}, y_{N2}), \cdots, (x_{Nm}, y_{Nm})\}$

(2) 共分散行列の計算

学習用データを用いて，攻撃データからなる共分散行列 V_1 と正常データからなる共分散行列 V_2 を計算

(3) 固有ベクトルの計算

V_1 の最大固有値に対応する固有ベクトル e_1 と V_2 の最大固有値に対応する固有ベクトル e_2 を計算

(4) 主成分分析による直線の導出

攻撃データの平均点 $(\overline{x_A}, \overline{y_A})$ を通り，固有ベクトル e_1 の傾きをもつ直線の方程式 l_A と，正常データの平均点 $(\overline{x_N}, \overline{y_N})$ を通り，固有ベクトル e_2 の傾きをもつ直線の方程式 l_N を計算

(5) 分類器の学習

2 直線 l_A, l_N のなす角を 2 等分する直線の方程式 $y = px + q$ を計算し，学習用データの分類率が高い方の方程式を分類器として採用

3.2. 提案アルゴリズム 2

(1) 学習用データの準備

- 攻撃データ n 個
  $\{(x_{A1}, y_{A1}), (x_{A2}, y_{A2}), \cdots, (x_{An}, y_{An})\}$
- 正常データ m 個
  $\{(x_{N1}, y_{N1}), (x_{N2}, y_{N2}), \cdots, (x_{Nm}, y_{Nm})\}$

(2) 平均による学習用データの縮約

攻撃データの平均の座標 $(\overline{x_A}, \overline{y_A})$ と，正常データの平均の座標 $(\overline{x_N}, \overline{y_N})$ を計算

(3) 分類器の学習

2 点 $(\overline{x_A}, \overline{y_A})$，$(\overline{x_N}, \overline{y_N})$ の垂直二等分線を分類器として採用

4. 人工データによるシミュレーション評価

この章では，SVM と前章の提案手法を用いて SQL インジェクション攻撃を検出するシミュレーションを行う．シミュレーションを行うために，文献 [11][12] で用いたものと同様に，624 個の攻撃データと，Web ページのフォームに入力される文字列を想定して人工的に作成した 234 個の正常データを用意し，これらのデータの中から学習用データとして攻撃データ，正常データともに 20 個ずつランダムに抽出して試験用データを構成し，2 種類の提案アルゴリズムの評価を行った．

4.1. 提案アルゴリズム 1 の評価

(1) 試験用データの準備

- 攻撃データ s 個
  $\{(x_{A1}, y_{A1}), (x_{A2}, y_{A2}), \cdots, (x_{As}, y_{As})\}$
- 正常データ t 個
  $\{(x_{N1}, y_{N1}), (x_{N2}, y_{N2},), \cdots, (x_{Nt}, y_{Nt})\}$

(2) 主成分分析による直線の導出

s 個の攻撃データを直線 $l_A : y = p_A x + q_A$ に試験用攻撃データを (垂直方向に) 射影し，t 個の攻撃データを直線 $l_N : y = p_N x + q_N$ に試験用正常データを (垂直方向に) 射影

(3) 分類器による検出

- 攻撃データの検出
  $y_{Aj} \geq px_{Aj} + q (1 \leq j \leq s)$ ならば試験用攻撃データを攻撃データとして検出
  $y_{Aj} < px_{Aj} + q (1 \leq j \leq s)$ ならば試験用攻撃データを正常データとして検出
- 正常データの検出
  $y_{Ni} < px_{Ni} + q (1 \leq i \leq t)$ ならば試験用正常データを正常データとして検出
  $y_{Nj} \geq px_{Ni} + q (1 \leq i \leq t)$ ならば試験用正常データを攻撃データとして検出

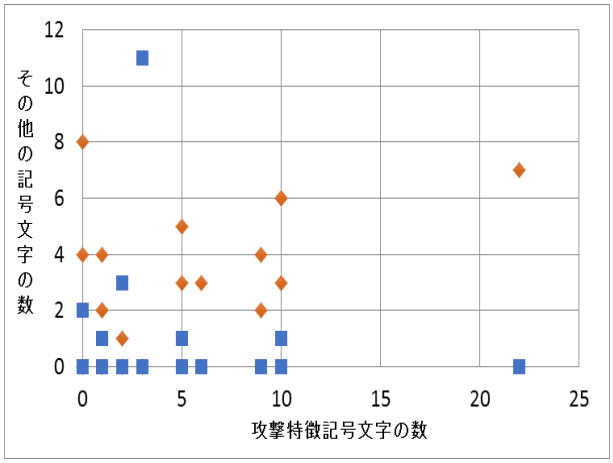

図 1: 学習用の攻撃データ（オレンジ）と正常データ（青）の散布図

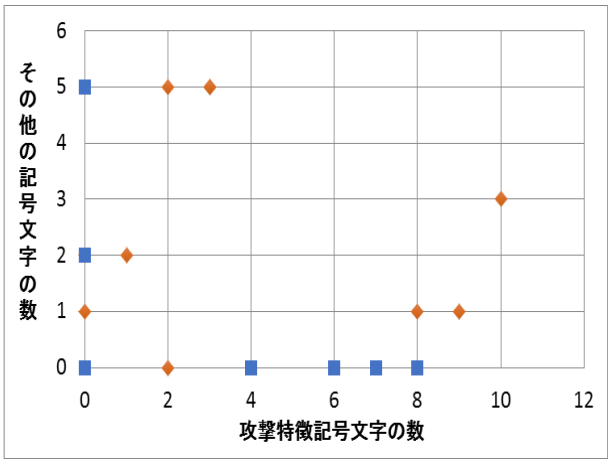

図 2: 試験用の攻撃データ（オレンジ）と正常データ（青）の散布図

4.2. 提案アルゴリズム 2 の評価

(1) 試験用データの準備

- 攻撃データ s 個
  $\{(x_{A1}, y_{A1}), (x_{A2}, y_{A2}), \cdots, (x_{As}, y_{As})\}$
- 正常データ t 個
  $\{(x_{N1}, y_{N1}), (x_{N2}, y_{N2}), \cdots, (x_{Nt}, y_{Nt})\}$

(2) 平均による分類器による検出分類器の直線の方程式が $y = ux + v$ の場合，

- 攻撃データの検出
  $y_{Aj} \geq ux_{Aj} + v(1 \leq j \leq s)$ ならば試験用攻撃データを攻撃データとして検出
  $y_{Aj} < ux_{Aj} + v(1 \leq j \leq s)$ ならば試験用攻撃データを正常データとして検出
- 正常データの検出
  $y_{Ni} < ux_{Ni} + v(1 \leq i \leq t)$ ならば試験用正常データを正常データとして検出
  $y_{Nj} \geq ux_{Ni} + v(1 \leq i \leq t)$ ならば試験用正常データを攻撃データとして検出

4.3.SVM による攻撃検出シミュレーション

SVM の計算には R version 2.15.3 を使用し，kernlab の ksvm 関数をカーネル関数を適用せずに利用した．図 1 に学習用データの散布図を示す．図 2 に試験用データの散布図を示す．R によるシミュレーションの結果，攻撃の試験用データに対しては 10 個中 2 個のデータについて正常と誤検出し，正常の試験用データについては 10 個中 6 個のデータについて攻撃と誤検出した．結果をまとめると適正検出率は 60% であった．ksvm 関数についてはカーネル関数にガウシアンを利用し，パラメータの値を 0.7 とした場合でも，カーネル関数を使わない場合と同様に適正検出率は 60.0% となった．

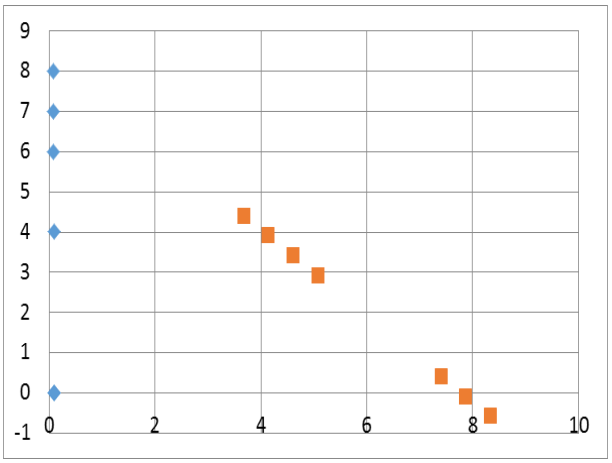

図 3: 射影された試験用の攻撃データ（オレンジ）と正常データ（青）の散布図

4.4. 提案アルゴリズム 1 による攻撃検出シミュレーション

学習用データの特徴空間の散布図は図 1 の通りである．これらのデータにアルゴリズム 1 を適用すると，直線 l_A, l_N の方程式は次のようになる．

$$l_A : y = -1.071x + 8.355, \quad (1)$$
$$l_N : y = -454.056x + 46.456. \quad (2)$$

直線 l_A, l_N 上に図 2 の試験用データを射影させると図 3 のようになる．また，直線 l_A, l_N の方程式から分類器の方程式は $y = -2.53x + 8.478$ となる．この分類器を用いて試験用データの分類を行ったところ，誤検出は 1 つも無く，適正検出率は 100.0% であった．

4.5. 提案アルゴリズム 2 による攻撃検出シミュレーション

学習用データの平均座標を計算すると攻撃データでは $(5, 3)$，正常データでは $(0.10, 1.05)$ となり，これら

の 2 点を通る直線の垂直 2 等分線を分離平面とする分類器 F の方程式は，

$$y = -2.513x + 8.433, \tag{3}$$

で与えられる．いま，正常の試験データを (x_N, y_N) とする．分類器 F は，

$$y_N < -2.513x_N + 8.433, \tag{4}$$

であれば，正常の試験データを正常と適正検知し，

$$y_N \geq -2.513x_N + 8.433, \tag{5}$$

であれば，F は正常の試験データを攻撃として誤検知する．実験の結果，$(5, 0)$ に対応する正常の試験データのみ誤検知したが，その他のデータはすべて適正検知した．また，攻撃の試験データを (x_A, y_A) とする．F は，

$$y_A \geq -2.513x_A + 8.433, \tag{6}$$

であれば，攻撃の試験データを攻撃と検知し，

$$y_A < -2.513x_A + 8.433, \tag{7}$$

であれば，攻撃の試験データを正常と誤検知する．実験の結果，$(0, 2), (1, 2), (2, 0), (0, 1)$ に対応する攻撃の試験データは誤検知し，他の試験データは適正検知した．以上より，F の適正検出率は 75.0% となった．

5. 考察とまとめ

本研究では，主成分分析を用いた分類器を構成する方法を提案し，SQL インジェクション攻撃のデータに対して提案法と SVM を適用し，攻撃を検出するシミュレーションを行った．この章では，4 章で得られたシミュレーションの結果について，特徴空間と分類器の構成法に関する観点から考察を行う．

5.1. データの特徴空間について

第 3 章で提案したアルゴリズムでは，入力文字列に含まれる攻撃特徴文字の出現頻度をデータの特徴空間として攻撃を検出する．攻撃特徴文字として使用した 5 つの文字は SQL インジェクション攻撃の文字列に頻出する傾向にあることから，著者らはこれらの文字をうまく組み合わせて文字列長で攻撃特徴文字の出現頻度の値を割って求められる攻撃特徴文字の含有率を用いて効率よく SQL インジェクション攻撃を検出できることを示している [12]．著者らの従来研究 [12] では，SQL の文法としてよく使われる 20 個の記号を用意して 5 文字の攻撃特徴文字を抽出したが，残りの 15 文字の記号は攻撃検出に使用しなかった．しかしながら，攻撃特徴文字としている 5 文字と残りの 15 文字の関係を調べることでより SQL インジェクション攻撃の特徴を捉えることができると考え，第 1 成分を攻撃特徴 5 文字の出現頻度，第 2 成分をその他の 15 文字の出現頻度とする 2 次元ベクトルをデータの特徴空間とし，この特徴空間に対して攻撃と正常それぞれのデータ群に対して主成分分析を適用した．

5.2. 分類器の構成法

第 4 章の図 1 で，分類器を構成するための学習用データの散布図を紹介した．この学習用データの攻撃データ群と正常データ群のそれぞれに対して主成分分析を適用したところ，攻撃データ群の第 1 主成分の固有ベクトルと平行な直線の傾き t_A と，正常データ群の第 1 主成分の固有ベクトルと平行な直線の傾き t_B は，以下の性質をもつことが確認された．

- t_A:絶対値の小さな負の実数
- t_B:絶対値の大きな負の実数

2 次元ベクトルの第 1 成分，つまり横軸には攻撃特徴文字の出現頻度を使用しているため，直線の傾きが絶対値の小さな負の実数であることは，攻撃文字列には攻撃特徴文字がたくさん含まれていることを示しているだけでなく，攻撃特徴文字の出現頻度が多くなってもその他の文字の出現頻度が多くなる傾向にないことが示されていると考えられる．このことはまさに攻撃特徴文字の 5 文字は文字列長に現れる特徴となっていることを表しており，他の 15 文字は攻撃の特徴としては上手く働いていないことを示している．

一方，直線の傾きが絶対値の大きな負の実数であることは，正常文字列には攻撃特徴文字が含まれることは少なく，その他の 15 文字の出現頻度は攻撃特徴文字の出現頻度より大きくなる傾向があることを示していると考えられる．したがって，正常文字列に現れる攻撃特徴 5 文字とその他の 15 文字の割合を見ることで，入力文字列が正常文字列であるかどうかを検出できる可能性があると考えられる．しかしながら，正常文字列には文字列として意味をなさない記号の列も含まれるため，攻撃特徴文字 5 文字とその他の 15 文字の割合によって，正常文字列を検出できるとは限らないという問題を含んでいる．

以上の考察から，主成分分析を使用して分類器を構成する提案アルゴリズム 1 では，攻撃データ群の平均の座標を通り，傾きを t_A とする直線を求め，攻撃データを垂直方向にその直線に射影することで攻撃を検出し，正常データ群にも同様の方法で検出する手法を採用した．

提案アルゴリズム 2 は主成分分析を用いず，単純に攻撃データ群と正常データ群の平均の座標をデータの代表点とし，2 つの代表点を結ぶ直線の垂直二等分線を分類器としてシンプルに攻撃を検出する方法として考えた．

参考文献

[1] NT Web Technology Vulnerabilities [online], http://www.phrack.com/issues.html?issue=54

[2] Vladimir N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, 1998.

[3] 伊波靖, 安里梓, 高良富夫,「SVM を利用した WAF の検知手法の提案」, 情報処理学会全国大会講演論文集 2011, pp. 3-445–3-447, 2011 年 3 月.

[4] Romil Rawat, Shailendra Kumar Shrivastav, “SQL injection attack Detection using SVM,” International Journal of Computer Applications, Volume 42, No.13, pp. 1–4, March 2012.

[5] Yi Wang, Zhoujun Li, “SQL Injection Detection with Composite Kernel in Support Vector Machine,” International Journal of Security and Its Applications, Vol. 6, No. 2,pp. 191–196, April 2012.

[6] Fuzhi ZHANG, Dongyan JIA, Jinbo CHAO, “An Improved PCA Attack Detection Algorithm Based on Normal Cloud Model,” Journal of Computational Information Systems 6:6(2010) 1959–1966, 2010.

[7] Gholam Reza Zargar, Tania Baghaie, “Category-Based Intrusion Detection Using PCA,” Journal of Information Security, 2012, 3, 259-271, 2012.

[8] DARPA Intrusion Detection Data Sets [online], http://www.ll.mit.edu/mission/communications/cyber/CSTcorpora/ideval/data/

[9] KDD Cup 1999 Data [online], http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[10] The NSL-KDD Data Set [online], http://nsl.cs.unb.ca/NSL-KDD/

[11] 園田道夫, 松田健, 小泉大城, 平澤茂一, 「文字単位の特徴抽出による SQL インジェクション攻撃検出法について」, 情報処理学会研究報告（コンピュータセキュリティ）, vol. 49, pp. 1–7, 2011 年 3 月.

[12] Michio Sonoda, Takeshi Matsuda, Daiki Koizumi, and Shigeichi Hirasawa, “On Automatic Detection of SQL Injection Attacks by the Feature Extraction of the Single Character,” Proceedings of the 4th International Conference on Security of Information and Networks (SIN2011), pp. 81–86, Nov. 2011.