5E-3

文字集合からの特徴抽出による SQLインジェクション攻撃の自動検出における閾値学習アルゴリズム

小泉　大城†　　松田　健†　　園田　道夫†　　平澤　茂一†

†サイバー大学 IT 総合学部

1　はじめに

SQL インジェクション攻撃の自動検出法としては，構文解析による方法やブラックリスト方式などがすでに実用化されている．しかし，近年の SQL インジェクション攻撃の種類の急激な増加にともない，リストの更新の処理コストや，検出の際の計算コストが肥大化し，対応が困難になってきている．そこで著者らは，攻撃を特徴づける文字が SQL データベースに入力された文字列の中に占める割合をもとに，SQL インジェクション攻撃の自動検出を行うアルゴリズムを提案した [1]. さらに，SQL データベースへの正常入力および SQL インジェクション攻撃入力の両方からなる人工データのサンプルを用意したもとで，この提案アルゴリズムの評価を行った [1]．本研究では，このアルゴリズムによる SQL インジェクション攻撃の自動検出にあたり，既知として必要な閾値を入力データのサンプルから学習するアルゴリズムを提案し，人工データのサンプルによりその性能を実験的に評価した結果を報告する．

2　SQL インジェクション攻撃の自動検出アルゴリズム [1]

2.1　準備

いま，Web アプリケーションなどを通じた SQL データベースへの入力を $l_i, i = 1,2,\cdots,n$ とする．ここでそれぞれの l_i は，正常入力または SQL インジェクション攻撃入力のどちらかである．l_i が正常入力であるか，(SQL インジェクション) 攻撃入力であるかを自動検出するにあたり，着目する既知の単一文字を $s_i, s_{ii}, s_{iii}, \cdots$ とする．単一文字には，たとえば半角スペース，セミコロン，シングルクォーテーションなどが該当するが，どの文字 (複数個も可) に着目するか，あらかじめ有限数決めておく．これらの単一文字 s のうち，空集合を含まないべき集合を $S_k, k = 1,2,\cdots,m$ とし，これを文字集合と呼ぶ．たとえば，着目する単一文字を半角スペース，セミコロンの 2 種類としたとき，文字集合の候補は，$S_1 = \{s_i\}, S_2 = \{s_{ii}\}, S_3 = \{s_i, s_{ii}\}$ の 3 通りとなる．

2.2　攻撃入力の自動検出アルゴリズム [1]

2.1 節の定義から，著者らは次のような SQL インジェクション攻撃の自動検出アルゴリズムを提案した [1].

1. 既知の値の事前設定
   (a) 文字集合の候補 $S_k, k = 1,2,\cdots,m$ の中から文字集合をひとつ設定.
   (b) 自動検出の閾値として用いる実数定数 $\alpha \in [0,1]$ を設定.
2. p_{ik} の計算
   入力文字列 l_i について，式 (1) により p_{ik} を計算する：
   $$p_{ik} = \frac{\#S_k}{l_i}. \tag{1}$$
3. 攻撃入力の自動検出
   入力文字列が正常入力 (0) であるか，攻撃入力 (1) であるかを式 (2) の関数 $h(p_{ik}, \alpha)$ により自動検出する：
   $$h(p_{ik}, \alpha) = \begin{cases} 0 & if\ \ p_{ik} \le \alpha; \\ 1 & if\ \ p_{ik} > \alpha. \end{cases} \tag{2}$$

2.3　攻撃の自動検出アルゴリズムの評価

2.2 節のアルゴリズムを評価するにあたり，624 種類の SQL インジェクション攻撃入力，および 234 種類の正常入力を考慮し，人工データを構成した．ここで，実用時の設定を考慮すると，SQL データベースへの入力が攻撃入力であるか，正常入力であるかは未知である．そこで，シミュレーションによる評価の際には，これらの人工データからサンプルデータを一定数，一様分布によりランダムに生成した．その際，$0 \le \beta \le 1$ なる実数 β をとり，攻撃入力と正常入力の総数が $(1-\beta) : \beta$ となるようにサンプリングした．

このもとで，正常入力をただしく検出した検出率を p_A, 攻撃入力をただしく検出できた検出率を p_N として，式 (3) によって総合的検出率 $\mu(S_k, \alpha, \beta)$ を計算した：

$$\mu(S_k, \alpha, \beta) = \beta p_N + (1-\beta) p_A. \tag{3}$$

以上により，自動検出アルゴリズムにおいて既知でなければならない文字集合 S_k および閾値 α と，評価基準である総合的検出率 $\mu(S_k, \alpha, \beta)$ との関係の考察がなされている [1].

On the Threshold Learning Algorithm of SQL Injection Attack Detection by the Feature of the Single Character

†Daiki KOIZUMI, Takeshi MATSUDA, Michio SONODA, and Shigeichi HIRASAWA are with Faculty of Information Technology and Business, Cyber University, Tokyo, JAPAN.

3 閾値学習アルゴリズム

3.1 閾値 α の学習規準

著者らの従来の研究 [1] では，閾値 α を次のような手順で経験的に決定する手法について扱った．すなわち，有限数の候補 α_j，$j = 1, 2, \cdots$ を用意し，文字集合 S_k，攻撃入力の比率 β が既知のもとで，総合検出率 $\mu(S_k, \alpha, \beta)$ が最大となるように α_j を経験的に決定する，というものである．

このようにして決定される閾値 α^* とすると，α^* は式 (4) のように定式化することができる．

$$\alpha^* = \arg\max_{\alpha_j} \mu(S_k, \alpha_j, \beta). \tag{4}$$

しかしながら，実用時においては入力に対する β の値は未知である．そこで，様々に変動する β の確率分布を考慮し，正常入力や攻撃入力についての教師付きサンプルデータが得られたもとで，閾値 α の学習規準を考える．

このとき，様々な入力に対して刻々と変動する β に対し，安定した総合検出率 μ を達成するような閾値 α が満たすべき要件として，以下の 2 項目が考えられる：

- $0 \le \beta \le 1$ なる β の区間に対して，$\mu(S_k, \alpha_j, \beta)$ の値がなるべく大きい
- $0 \le \beta \le 1$ なる β の区間に対して，$\mu(S_k, \alpha_j, \beta)$ の変動がなるべく少ない

上記の 2 項目を満たすような α を特に $\hat{\alpha}$ とすると，β の閉区間 $[0, 1]$ を合計 M 個の部分区間 $\beta_1, \beta_2, \cdots, \beta_M$ に分割したもとで，$\hat{\alpha}$ の学習規準は，式 (5) のように定式化できる．

$$\begin{aligned}\hat{\alpha} \quad = \quad & \arg\max_{\alpha_j} \Bigg[\sum_{m=1}^{M} \beta_m \mu(S_k, \alpha_j, \beta_m) \\ & - \left| \frac{1}{M \sum_{m=1}^{M} (\beta_m)^2 - \left(\sum_{m=1}^{M} \beta_m \right)^2} \right| \\ & \times \Bigg| M \sum_{m=1}^{M} \beta_m \mu(S_k, \alpha_j, \beta_m) \\ & - \left(\sum_{m=1}^{M} \beta_m \right) \left(\sum_{m=1}^{M} \mu(S_k, \alpha_j, \beta_m) \right) \Bigg| \Bigg]. \end{aligned} \tag{5}$$

式 (5) の右辺の第 1 項では，上記の 2 点の要件のうちの 1 つめの項目を変数 μ に関して β の分布でとった期待値の近似値として計算している．つづく第 2–4 項では，2 つめの項目を，M 個の説明変数と被説明変数の組 $\left(\beta_m, \mu(S_k, \alpha_j, \beta_m)\right)$ に対する単回帰直線の傾きの絶対値として計算している．式 (5) では，両者を等重みで考慮することにより，有限数の候補 α_j の中で極大となる閾値 $\hat{\alpha}$ が選択される．

3.2 提案アルゴリズム

式 (5) を用いると，以下のような閾値学習アルゴリズムが得られる．

1. 閾値 α の候補として，α_j，$j = 1, 2, \cdots$ を準備
2. β の閉区間 $[0, 1]$ を合計 M 個の部分区間 β_m，$m = 1, 2, \cdots, M$ に分割
3. α_j, β_m を既知として 2.2 節のアルゴリズムにより総合検出率 $\mu(S_k, \alpha_j, \beta_m)$ を計算
4. すべての α_j, β_m の組と前のステップで得られた $\mu(S_k, \alpha_j, \beta_m)$ とから，式 (5) の意味で最適な $\hat{\alpha}$ を選択

3.3 提案アルゴリズムの適用例

3.2 節に述べたアルゴリズムの適用結果の一例を図 1 に示す．図 1 では，α_j として，$\alpha_1 = 0.01, \cdots, \alpha_{12} = 0.12$ の計 12 種類の値を，β_m として，$\beta_1 = 0.05, \beta_2 = 0.10, \cdots, \beta_{20} = 1.00$ の計 20 種類の値を考慮し，計算を行った．式 (5) の値は，$\alpha_2 = 0.02$ のとき 0.758 に，$\alpha_9 = 0.09$ のとき 0.904 に，$\alpha_{11} = 0.11$ のとき 0.795 となり，全体としては α_9 で極大となった．

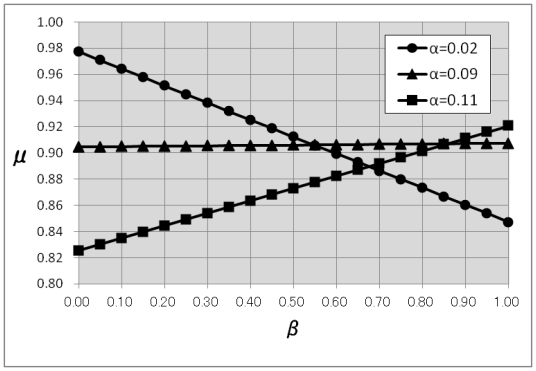

図 1: 様々な α における総合検出率 μ と β の関係

4 考察

図 1 からわかるように，提案アルゴリズムにおいて $\hat{\alpha} = 0.09$ においては，正常入力と攻撃入力の比率 β が変化しても安定した検出率 μ を実現していることがわかる．特定の β のみに着目して α を決定すると，例えば $\beta = 0.00$ のときに $\alpha^* = 0.02$ を選ぶような事態が起きうるが，このときは正常入力の検出には適するが，攻撃入力の検出には検出率が低下してしまう．

5 おわりに

本研究では，SQL インジェクション攻撃の自動検出に必要な既知の閾値を学習するアルゴリズムを提案した．また，その適用例を示し性能を確認した．今後の課題としては，正常入力のサンプルの充実や実データへの適用実験およびその評価などが挙げられる．

参考文献

[1] Michio SONODA, Takeshi MATSUDA, Daiki KOIZUMI, and Shigeichi HIRASAWA, “On Automatic Detection of SQL Injection Attacks by the Feature Extraction of the Single Character,” Proceeding of 4th International Conference on Security of Information and Networks (SIN2011), pp.81–86, Nov. 2011.