様式Ｆ－１９

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成２５年　６月１７日現在

機関番号：３２８１７
研究種目：若手研究（B）
研究期間：2011～2012
課題番号：２３７４００９４
研究課題名（和文）実特異点の解析とその情報数理への応用

研究課題名（英文）Analysis of real singularities and its application to mathematical information

研究代表者
松田　健（MATSUDA　TAKESHI）
サイバー大学・総合情報学部・講師
研究者番号：４０５９１１７８

研究成果の概要（和文）：
本研究は，ベイズ学習理論を情報セキュリティの分野へ応用するために，ある実代数的集合の特異点解析をすることで汎化誤差の理論値に関して考察し，Web アプリケーションの脆弱性を狙った攻撃の特徴を抽出して攻撃を自動検出するための確率モデルの開発手法について検討するものである．ベイズ学習理論を用いて Web アプリケーション攻撃を検出することの有用性を理論と実験の両面の側面から示した．

研究成果の概要（英文）：
In this study, we studied on real singularities analysis in Bayesian inference and its application to information security. Firstly, we studied on real singularities of Bayesian inference and the generalization error. Secondly, we studied on the method to apply Bayesian inference to the detection of web application attacks. In that process, we proposed the extraction method of attack features using symbols in strings. By using the feature extraction method, we proposed a new stochastic model to detect web application attacks. Finally, we showed that the effectiveness of Bayesian inference concerning the application of the detection of web application attacks.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
交付決定額	2,000,000	600,000	2,600,000

研究分野：数物系科学
科研費の分科・細目：数学・数学一般（含確率論・統計数学）
キーワード：ベイズ学習，実特異点，汎化誤差，Web アプリケーション攻撃，自動検出

１．研究開始当初の背景

ベイズ学習理論はベイズ理論を用いた統計的推測の一手法であり，情報工学や情報科学のみならず，医療情報・社会科学・経済・生物情報など多岐に渡る分野において利用され，その有用性が示されている．一方，上記の応用では一般的に複雑な確率モデルを用いて実現されるが，ベイズ学習を推測に用いた場合，確率モデルの推測の精度を示す指標としてベイズ汎化誤差が広く利用されている．しかし，正則でない確率モデルのベイズ汎化誤差を求めるためには実特異点集合の解析が必要になるため，任意の確率モデルのベイズ汎化誤差を求めることは容易ではなく，正則でない確率モデルの理論的な有用性はまだ十分に調べられているとは言えないのが実情である．また，複雑な確率モデルの汎化誤差が得られている場合でも，そのような確率モデルをどのように実際の問題に適用するかということについてはそれぞれの分野の問題毎に検証する必要がある．そこで以上のことを踏まえて，複雑な確率モデル

のパラメータ空間がもつ実特異点の解析をしてその性質からベイズ汎化誤差の性質を調べること，さらにそのような確率モデルにベイズ学習理論を適用して Web アプリケーション攻撃を自動検出する方法を開発することを目的として本研究を遂行した．統計的推測のセキュリティ分野への応用として最も有名なものはナイーブベイズの迷惑メールフィルタリングへの応用であるが，研究代表者の松田が本研究に取り組む以前からマルウェアや Web アプリケーション攻撃の自動検出に統計的推測の手法を用いた研究が存在していた．しかし，統計的推測や機械学習を用いた Web アプリケーション攻撃の自動検出については一定の成果はあるものの，それらの検出を回避する攻撃が開発されるという問題があり，これらの方法は今のところ有効な攻撃検出方法として確立されていないのが実情である．このような現状に基づき，本研究は Web アプリケーション攻撃を自動検出するための確率モデルを開発し，その有用性を検証するものである．

２．研究の目的

本研究の目的は，複雑な確率モデルの汎化誤差に関する情報を得るために確率モデルがもつパラメータ空間の実特異点の解析を行うことであり，そのような確率モデルが情報セキュリティの分野に応用可能かどうかを検証することである．確率モデルが正則でない場合のベイズ汎化誤差は正則である確率モデルと比べて小さくなることが知られているため，正則でない複雑な確率モデルにベイズ理論を適用して統計的推測を行うことの良さについては理論的に示されているといえる．しかしながら，実際に確率モデルの応用を考える際には確率モデルのパラメータに何らかの意味をもたせる必要もあるため，応用先のそれぞれの分野の特性に合わせながら確率モデルを検討しなければならない．そこで本研究では課題を以下の三点に絞り，研究を遂行した．

(1) 第一の課題

確率モデルのベイズ汎化誤差の導出の際に現れる特異点を定義する多項式の特徴について調べ，ベイズ汎化誤差との関連を調べる．

(2) 第二の課題

Web アプリケーション攻撃の自動検出が可能な確率モデルの開発をするために SQL インジェクション攻撃とクロスサイトスクリプティング攻撃のデータを収集し，解析する．

(3) 第三の課題

第二の課題の結果を踏まえて，SQL インジェクション攻撃とクロスサイトスクリプティング攻撃を自動検出するための正則でない確率モデルを検討する．

３．研究の方法

(1) 第一の課題

ベイズ汎化誤差を求めるには，ある実多項式の代数的集合の特異点の解消が必要となる．本研究では 2 つの異なる実多項式 $f, g \geqq 0$ の和として定義される実多項式 $f+g \geqq 0$ について考察し，いくつかの具体例を計算することで多項式 f，g，f+g から求められるベイズ汎化誤差について調べ，その性質を整理した．

(2) 第二の課題

Web アプリケーションの脆弱性を標的とする攻撃は様々なものがあるため，まず代表的な Web アプリケーション攻撃である SQL インジェクション攻撃について書籍や Web から攻撃データを収集し，分析した．その際，攻撃データでないデータの解析も必要となるが，一般的に攻撃でないデータには個人情報が含まれる場合もあるため，攻撃でないデータは住所・氏名などの一般的な個人情報の他，Wiki の文法や顔文字などのデータを収集してデータの解析をした．

SQL インジェクション攻撃を未然に防ぐ技術的対策としては，シングルクォートやセミコロンなどの記号入力を制限する方法が広く使われているが，この方法では入力制限した文字を使わない攻撃を検出することができない．しかし，SQL インジェクション攻撃はいくつかの特殊な記号を含むことが多いため，本研究では SQL の文法で利用される記号に着目し，記号の出現頻度を解析することで SQL インジェクション攻撃の特徴となりえる記号を抽出した．さらにその解析結果をもとにして攻撃を検出するための確率モデルを提案し，その有用性を理論的に示した．また，クロスサイトスクリプティング攻撃についても同様の検討を行った．

(3) 第三の課題

SQL インジェクション攻撃とクロスサイトスクリプティング攻撃のデータ（以下，攻撃データと呼ぶ）を解析した際に，攻撃でない通常の SQL クエリ（以下，正常データと呼ぶ）の分析も行った．その際に，攻撃・正常データの文字列に含まれる攻撃特徴となる記号（以下，攻撃特徴文字）の含有率を攻撃データ群と正常データ群で比較したところ，攻撃特徴文字の含有率をもとにして定義できる混合分布を用いて攻撃・正常データの性質を表すことができた．

４．研究成果

(1)ベイズ汎化誤差については 2 つの実多項式の和で定義される多項式の代数的集合の特異点を考え，その上界と下界がそれぞれの実多項式の代数的集合の特異点解消から得られるための 1 つの条件を与えた．また，混

合数が 2 である混合 4 項分布の特異点を解消し，そのベイズ汎化誤差の漸近展開の主要項の係数を導出した．

(2)収集した SQL インジェクション攻撃データについて SQL 文によく利用される 20 個の記号を攻撃特徴文字の候補としてそれらの出現頻度を調べ，各攻撃データにそれぞれの記号が含まれている割合（以下，記号の含有率と呼ぶ）を調べたところ表 1 のような結果が得られた．

表 1　攻撃特徴文字候補と検出率

攻撃特徴文字候補	P_A	P_N
s_{i}（スペース）	0.971	0.098
s_{ii}（セミコロン）	0.664	0.000
s_{iii}（シングルクォーテーション）	0.483	0.000
s_{iv}（左側丸括弧）	0.459	0.060
s_{v}（右側丸括弧）	0.414	0.090
s_{vi}（左側中括弧）	0.000	0.060
s_{vii}（右側中括弧）	0.000	0.026
s_{viii}（左側大括弧）	0.000	0.026
s_{ix}（右側代括弧）	0.000	0.026
s_{x}（シャープ）	0.032	0.021
s_{xi}（パーセント）	0.024	0.026
s_{xii}（ダブルクォーテーション）	0.010	0.000
s_{xiii}（アンパサンド）	0.019	0.021
s_{xiv}（バックスラッシュ）	0.032	0.000
s_{xv}（パイプ）	0.040	0.103
s_{xvi}（等号）	0.294	0.060
s_{xvii}（大なり記号）	0.000	0.090
s_{xviii}（小なり記号）	0.000	0.038
s_{xix}（アスタリスク）	0.128	0.094
s_{xx}（スラッシュ）	0.112	0.073

表 1 において，記号 s_a (a=i, ii, …, xx)の含有率が 0.1 以上であるものを攻撃データとして検出した場合，P_Aは攻撃データを攻撃と正しく検出した割合，P_Nは正常データを攻撃と誤って検出した割合を示すものである．表 1 からは，例えばスペースを攻撃特徴としてその含有率が 0.1 以上のデータを攻撃とみなして SQL インジェクション攻撃を検出した場合，収集した攻撃データの約 97%のデータを攻撃と検出し，正常データの約 10%を攻撃と誤検出したことが分かる．また，SQL インジェクション攻撃によく利用されるシングルクォートやセミコロンなどの記号についても，これらの記号を必要としない攻撃もあるために P_A の値はそれほど高い値にはならなかった．そこで，表 1 の結果をもとに，いくつかの攻撃特徴文字の含有率をもとに攻撃検出の実験を行ったところ収集したデータにおいては{スペース，セミコロン，ダブルクォート，右側丸括弧，左側丸括弧}の 5 つの攻撃特徴文字で検出する方法が効果的であるという結果が得られた．また，それらの文字の含有率の値からそのデータが攻撃データである確率を表現する関数を定義し，それをベルヌーイ分布に応用した確率モデルを提案し，2 値分類で広く利用されているシグモイド関数を応用した確率モデルとの予測誤差の比較を行い，提案モデルの予測誤差がシグモイド関数を用いた確率モデルの予測誤差よりも小さくなることを数学的に示し，数値実験で具体的にどの程度の予測誤差の差が出るかということを示した(図 1)．

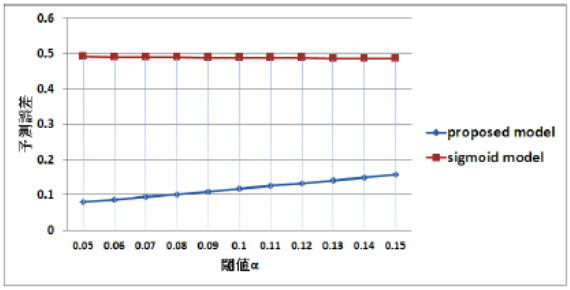

図 1　予測誤差の比較

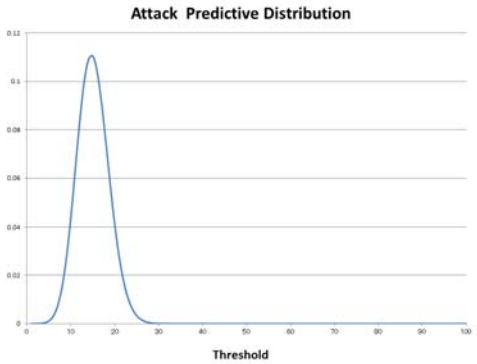

図 2　攻撃データの分布

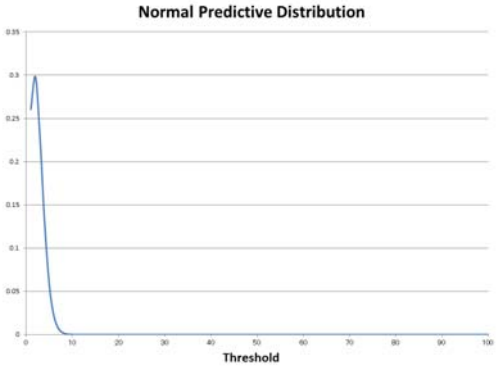

図 3　正常データの分布

さらに提案モデルとシグモイド関数を用いたモデルに最尤推定法を適用し，ニュートンラプソン法によって最尤推定量を計算する

際に，提案モデルの方が収束までの繰り返し回数が少なくなることを実験により示した．また，上述の提案方法を 2 項分布に応用した確率モデルを提案し，それにベイズ学習を適用して攻撃データと正常データの分析をしたところ攻撃データと正常データでは図２，３に示す通り，データの分布がまったく異なるものになることを確認することができた．図２，３の横軸は記号の含有率を 100 倍した数の整数部分を表している．図２は攻撃データを，図３は正常データを二項分布で表したものである．攻撃データでは分布のピークが横軸の 15～20 付近で，正常データではピークが 0～5 付近で現れることが大きな特徴となっており，この特徴を用いて SQL インジェクション攻撃を検出する手法を提案した．

一方，SQL インジェクション攻撃の検出に利用した攻撃特徴文字の含有率を用いる手法をクロスサイトスクリプティング攻撃の

表２　クロスサイトスクリプティング攻劇の特徴

Variable	Candidates of Symbols	Attack Feature Value
s_1	”	7.711632
s_2	>	6.251653
s_3	/	5.799894
s_4	<	5.464814
s_5	space	3.426871
s_6	=	2.337674
s_7	’	2.190181
s_8	:	1.218685
s_9	.	1.203845
s_{10}	(	1.133619
s_{11}	)	1.020919
s_{12}	-	0.770402
s_{13}	;	0.652083
s_{14}		0.200379
s_{15}	&	0.150932
s_{16}	{	0.138507
s_{17}	}	0.130828
s_{18}	#	0.118164
s_{19}	+	0.116853
s_{20}	!	0.106243
s_{21}	,	0.035590
s_{22}	@	0.028701
s_{23}	?	0.027731
s_{24}	]	0.02139
s_{25}	[	0.021167
s_{26}	–	0.007839
s_{27}		0.007763
s_{28}	*	0.007723
s_{29}	|	0.007676
s_{30}	^ (caret)	0.007489
s_{31}	%	0.007489
s_{32}	$	0.007412

データに適用し，表１と同様な考察を行ったところ，P_A, P_N どちらの値も 0.1 以下になり，SQL インジェクション攻撃の検出のときと同様に記号を組み合わせて検討しても，顕著な特徴を得ることは出来なかった．これはクロスサイトスクリプティング攻撃が html のタグや javascript の関数で構成されているため，攻撃の文字列長が長くなり，記号の含有率は小さくなるからだと考えられる．そこで html のタグの性質を考え，クロスサイトスクリプティング攻撃に含まれる記号の出現頻度だけでなく，出現位置を加味してそれぞれの記号にスコア付け（表２の Attack Feature Value）してそれを用いて攻撃を検出する手法を提案したところ，図４のようにクロスサイトスクリプティングの攻撃データと正常データを分けることができ，Attack Feature Value の値が 15 以上のデータを攻撃として検出したところ，99.5%の攻撃データと 97.5%の正常データを正しく検出することが出来た．ただし，学習用に利用したデータは（攻撃データ：70 サンプル，正常データ：30 サンプル）であり，テスト用には学習用データとは異なるデータ（攻撃：307 サンプル，正常:210 サンプル）を用意して実験を行った．テスト用のデータにはそれぞれのデータに番号のラベルをつけており，図４の横軸はそのラベルの番号を表し，それに対応する縦軸の値をもとにして攻撃の検出を行った．さらに，攻撃を検出するための閾値は学習データの違いによってある程度のばらつきをもつため，確率微分方程式を応用して攻撃検出を行う手法も提案し，その有用性を示した．

(3) 上記の研究成果から SQL インジェクション攻撃やクロスサイトスクリプティング攻撃といった Web アプリケーション攻撃においては，攻撃データと正常データをデータに含まれる記号の特徴を用いて分布の異なる確率モデルで表すことが可能であるということが分かった．このような性質から SQL インジェクション攻撃やクロスサイトスクリプティング攻撃を検出する混合数が 2 個の混合 4 項分布を提案し，その特異点を解消してベイズ汎化誤差の漸近展開の主要項の係数を導出した．しかし，実際に攻撃検出に使用するモデルは混合 4 項分布をさらに改良したものとなるため，そのようなモデルのベイズ汎化誤差について調べることが今後の課題である．

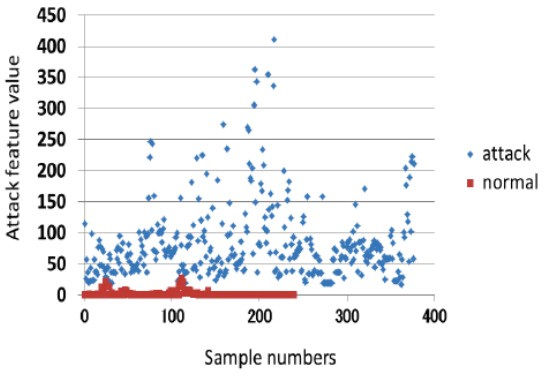

図４　クロスサイトスクリプティング攻撃の検出

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計９件）

① Takeshi Matsuda,
Desingularization of Mixture Tetranomial Distributions and Its Application for the Detection of Web Application Attacks,
The Proceedings of ICMA 2012 (ISSN1224-6069), 107-112, 2013, 査読有

② 松田健,
SQL インジェクション攻撃自動検出支援モデルと予測誤差,
情報処理学会論文誌 数理モデル化と応用, Vol.6, No.1, 10-19, 2013, 査読有

③ Takeshi Matsuda, Daiki Koizumi, Michio Sonoda,
Cross Site Scripting Attacks Detection Algorithm Based on the Appearance Position of Characters,
Proceedings in the 5th International Conference on Communications, Computers and Applications,
53-58, 2012, 査読有

④ Takeshi Matsuda,
On the classification based on the stochastic method,
Mathematical Sciences International Research Journal, Volume 1, Number 2, 394-402, 2012, 査読有

⑤ Takeshi Matsuda,
Estimation of SQL Injection Attacks Based on Single Character,
The 2012 International Conference on Applied and Theoretical Information Systems Research, CD-ROM, 2012,
査読有

⑥ Takeshi Matsuda, Daiki Koizumi, Michio Sonoda, Shigeichi Hirasawa,
On Predictive Errors of SQL Injection Attack Detection by the Feature of the Single Character,
Proceedings of The 2011 IEEE International Conference on Systems, Man, and Cybernetics, 1722-1727,
2011, 査読有

⑦ Takeshi Matsuda,
On Upper and Lower Boundaries of Real Log Canonical Threshold and Free Energy,
The 7th International Conference on Mathematics, Statistics and its Applications (ICMSA 2011), pp.226-235,
2011, 査読有

〔学会発表〕（計６件）

① Takeshi Matsuda,
Desingularization of Mixture Tetranomial Distributions and Its Application for the Detection of Web Application Attacks,
The 13th International Conference on Mathematics and Applications,
2012 年 11 月 2 日,
University “Politehnica” of Timisoara, Romania

② Takeshi Matsuda, Daiki Koizumi, Michio Sonoda, Shigeichi Hirasawa,
Predictive distribution of SQL Injection attacks Detection Model,
11th World Meeting of the International Society for Bayesian Analysis 2012, 2012 年 6 月 28 日,
京都テルサ

③ 園田道夫, 松田 健, 小泉大城, 平澤茂一, 辻井重男,
攻撃文字列の特徴抽出と Web アプリケーションの自動検出へのアプローチ,
情報処理学会第 74 回全国大会講演論文集, 3, 557-558, 2012 年 3 月 8 日, 名古屋工業大学

④ 松田 健,
SQL インジェクション攻撃自動検出支援モデルと予測誤差,
情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告
2012-MPS-87(14), 1-6, 2012 年 3 月 2 日,
鹿児島県指宿市民会館

〔その他〕
ホームページ等
http://matsudalab.office-server.co.jp/linear_algebra/index.html

６．研究組織
(1)研究代表者
松田　健（MATSUDA TAKESHI）
サイバー大学・総合情報学部・講師
研究者番号：４０５９１１７８

(2)研究分担者
なし

(3)連携研究者
なし

(4)研究協力者
園田　道夫（SONODA MICHIO）
サイバー大学・総合情報学部・准教授