P-6

ベイジアンフィルタを用いた未知コンピュータウイルスの検出

Detection of unknown computer viruses using Bayesian filtering

○森谷和徳[1]，馬場彰太郎[2]

*Kazunori Moriya[1], Shotaro Baba[2]

Abstract:　Many new computer viruses have appeared day by day. In particular, the viruses which have not been previously seen are really problematic. In such a case, we cannot conduct a simple pattern matching to detect the viruses, as we generally have no information about unknown computer viruses. For this reason, we need to avoid somehow the situation of computer viruses spread in the world. In this article, we suggest a method that detects unknown computer virus by means of a Bayesian filtering approach.

１．概要

今日，コンピュータウイルスは日々増加している．その中で最も脅威なものは未知コンピュータウイルスである．未知のコンピュータウイルスは通常，シグネチャがデータベースに登録されていないため，単純なパターンマッチングによりその検出を行うことができない．そのため対策が施される前に世界中に蔓延してしまう恐れがある．本論文ではベイジアンフィルタを用いて，未知コンピュータウイルスを検出する手法を提案する．

２．擬似ウイルスプログラムの作成

本研究では，擬似ウイルスとして Java 言語を用いて，指定したディレクトリのファイルを全て削除してしまうようなプログラムを 50 個用意した．各擬似ウイルスは，ループの条件及びダミープログラムの追加などのプログラムの一部分を変えて互いに亜種となるプログラムである．また，本研究ではこれら各亜種のプログラムを未知ウイルスと仮定した．さらに，未知ウイルスかどうかの判定を行うため，これらの擬似ウイルスプログラムとは別に，インターネット上から無作為に選んだプログラム 20 個及び上記とは別に用意されたプログラム 30 個の合計 50 個のプログラムを非ウイルスとして扱った．

３．ウイルスの検出

３．１ 学習

本研究では，Paul Graham Bayes にちなんだ確率モデルであるベイジアンフィルタ[1]を用いてウイルス及び非ウイルスの実行可能ファイルから表示可能文字列を抽出し学習を行った．また，表示可能文字列の長さは 4 文字以上とした．これは，3 文字以下の表示可能文字列を用いることで，多くの情報を得る事は可能となるが，ウイルス判定において余分な情報が混ざることで学習やカテゴリ分類する時間が長くなる恐れがあり，逆に，文字数の多い表示可能文字列では，ウイルス判定において必要な情報切り捨てられる恐れがあるためである[2]．この過程で，抽出された 4 文字以上の表示可能文字列を特徴点 s とし，各特徴点 s のウイルス確率 P(s)を求める．この時，ウイルスファイルの総数を n_{virus}，非ウイルスファイルの総数を n_{normal}，ある特徴点 s がウイルスに表れた回数を b，非ウイルスファイルに表れた回数を g とすると，ある特徴点のウイルス確率 p(s)は

$$p(s)=\frac{\frac{b}{n_{virus}}}{\frac{2g}{n_{normal}}+\frac{b}{n_{virus}}} \qquad (1)$$

と表すことができる．

ただし，ウイルス確率 p(s)は最大値を 0.99，また，最小値を 0.01 とする．ウイルスファイル及び非ウイルスファイルの総数を増やすことで，ウイルス確率 p(s)の，最大値 0.99 及び最小値 0.01 に近づくため，この値は経験的に得られた最適な値とされている[1]．なお，本研究では，2g+b<5 であるような出現回数が少ない特徴点はファイル固有の特徴点であるので，これに該当する場合は無視して計算を行わないこととした[2]．

以上により，ウイルスにのみしか表れない特徴点は 0.99，非ウイルスにしか表れない特徴点は 0.01 が与えられ，ウイルス及び非ウイルスに表れる特徴点は，表れる回数によって適切なウイルス確率が与えられる．ウイルスと非ウイルスに含まれる 2g+b<5 となる特徴点を除くすべての特徴点に対し，(1)式による計算を行うことで，特徴点のウイルス確率が収められたデータベースが作成できる．

３．２ ウイルスの検出

次に，ウイルス検出について述べる．以下は，学習

１：日大理工・院（前）・数学　２：日大理工・研究生・数学

後にある対象のプログラムが入力されウイルス判定されるまでの流れである:

（1）対象のプログラムから特徴点を抽出し，すべての特徴点に対しデータベースを参照しウイルス確率を割り当てる．なお, データベース中に対応する特徴点が存在しない場合は非ウイルス及びウイルスの特徴を示すものではなくファイル固有の特徴点と解釈し，やや無害な 0.4 を割り当てるものとした.

（2） 特徴点の確率が 0.5 から差が大きい順に特徴点を 15 個選択して後のウイルス判定に用いる．0.5 から差をとる理由は，対象のプログラムの性質を強く示す特徴点を選択するためである.

（3）（2）で選択された 15 個の複合確率を求めることでファイル全体のウイルス確率が求められるので, 選択された特徴点のウイルス確率を P_i とするとファイル全体のウイルス確率 P_v は

$$p_v = \frac{\prod_1^{i=15} P_i}{\prod_1^{i=15} P_i + \prod_1^{i=15} (1 - P_i)} \qquad (2)$$

と表すことができる.

ここで, 15 個の特徴点だけをウイルス判定に用いる理由を述べる: もし仮に全ての特徴点を(2)式に用いた場合，ウイルス制作者は無害なプログラムを大量に追加することで，容易に p_v 値を低下する事が出来る．一方, 15 個の特徴を優先するアルゴリズムであれば，ウイルス制作者は検出回避が困難になる.

（4）(2)式による p_v が設定したしきい値を超えた際にはウイルスとしてみなす.

4．実験

本研究におけるウイルス検出プログラムは, 以下のような手順により作成した:

（1） 50 個のウイルス及び同数個の非ウイルスからランダムに 20 個ずつ選んで計 40 個を学習データとして用い，残りの 30 個ずつ計 60 個をテストデータとして用いる.

（2）40 個の学習データから特徴点を抽出し各々の特徴点からウイルス確率を(1)式を用いて求め，特徴点のウイルス確率をデータベース化する.

（3）60 個のテストデータから特徴点を抽出し，ファイルのウイルス確率を求める．本研究では，しきい値を 0.9 と設定し 0.9 を超えた場合はウイルスと判定した. なお，ウイルスファイルのウイルス確率が 0.9 以下及び非ウイルスのウイルス確率が 0.9 を超えた場合に誤認識と判定した. また, 本研究ではしきい値を 0.9 と設定したが, 0.9 に近い値でも同様の結果を得ることができる．それは，p_v の値が性質上，極端に 1 もしくは 0 に近い値をとるためである．テストデータのウイルス確率の代表例を Table 1 に示す.

5．結論

4 で述べたプログラムを 50 回実行する．このプログラムは, ウイルス及び非ウイルス 50 個ずつのうちランダムに学習データ 20 個ずつ，テストデータ 30 個ずつに分割するプログラムなので，ウイルス検出率は毎回変化する．そこで，それぞれの実験における検出率の平均をとった．その結果，平均 97 パーセントの検出率を得た．しかし今回実験に扱ったウイルスは，検出が困難となるようなステルス技術については考慮しなかったため，比較的検出し易く,その結果, 比較的高い検出率が得られたと考えられる．今後はステルス技術が用いられたウイルスでも認識できる方法を検討することが課題である.

6．参考文献

[1] Paul Graham, "A Plan For Spam" , http://www.paulgraham.com/spam.html , August 2002.

[2] 小池竜一，中谷直司，萩原由香里，厚井祐司，高倉弘喜，吉田等明「ベイズアルゴリズムを用いた未知のコンピュータウイルス検出」, 情報処理学会論文誌, Vol. 46，No. 8，pp. 1986-1988，August 2005.

Table 1 Probabilities of virus infection

virusfile	ウイルス確率
virus24.class	0.999999997838189
virus9.class	0.999999997838189
virus14.class	0.999999999994926
virus30.class	0.999999999943849
virus45.class	0.999999997722040
virus33.class	0.999999997722040
virus42.class	0.999999999996165
virus41.class	0.999999999999983
virus2.class	0.999999997838189
virus19.class	0.999999999912654
novirusfile	ウイルス確率
novirus36.class	3.51E-12
novirus47.class	1.18E-22
novirus35.class	7.09E-16
novirus4.class	5.05E-09
novirus49.class	2.89E-19
novirus46.class	9.53E-11
novirus31.class	1.92E-13
novirus11.class	0.999999998
novirus43.class	1.18E-22
novirus32.class	7.09E-16