УДК 681.3.06

С.В. Поликарпов

ИССЛЕДОВАНИЕ НОВОЙ МОДЕЛИ ИСКУССТВЕННОГО НЕЙРОНА КАК ЭЛЕМЕНТА СИСТЕМЫ ОБНАРУЖЕНИЯ СЕТЕВЫХ АТАК

Рассматривается возможность применения новой модели искусственного нейрона в качестве элемента классификации системы обнаружения атак. Приводится краткое описание новой модели искусственного нейрона, основанной на использовании табличных подстановок вместо операций перемножения на весовые коэффициенты и нелинейной выходной функции. Предложен метод синтеза нейроклассификаторов в виде древовидных структур, основанный на поэтапном наращивании нейросети. На основе набора данных KDD cup'99 оценивается эффективность полученных нейроклассификаторов в сравнении с известными мировыми аналогами.

Искусственная нейросеть; обнаружение сетевых атак; классификация.

S.V. Polikarpov

INVESTIGATION OF A NEW MODEL OF ARTIFICIAL NEURON AS AN ELEMENT OF INTRUSION DETECTION SYSTEM

The possibility of a new model of artificial neuron as an element of the classification of intrusion detection systems. A brief description of a new model of artificial neuron, based on the use of table lookup operations, instead of multiplying the weighting coefficients and the nonlinear output function. Method is proposed for synthesizing neuroclassifiers in the form of tree structures, based on the gradual build a neural network. Based on the data set KDD cup'99 evaluated the effectiveness of the obtained neuroclassifiers in comparison with the known world analogues.

An artificial neural network; Intrusion detect systems; classification.

Актуальность разработки и исследования методов классификации. Одной из наиболее актуальных задач в эпоху развития распределённых информационных технологий является информационная безопасность глобальных и локальных сетей, в частности обнаружение сетевых атак в компьютерных сетях. Сложность данной задачи заключается в практическом отсутствии однозначного описания всех возможных сетевых атак и, как следствие, в невозможности их гарантированного обнаружения. Это объясняется такими факторами, как:

- наличие неизвестного количества уязвимостей программного обеспечения на сетевых узлах и сетевого оборудования;
- наличие уязвимостей в сетевых протоколах;
- ошибки в конфигурации сетевого оборудования.

Существует две актуальные проблемы, решение которых позволит значительно повысить эффективность систем обнаружения сетевых атак (IDS):

1) проблема автоматизации процесса выделения из сетевых атак специфических признаков;

2) проблема обобщения выделяемых признаков для обеспечения возможности обнаружения модификаций сетевых атак.

Для решения такого рода проблем в настоящее время активно развивается направление «data mining» (извлечение данных), в основе которого лежит развитие методов классификации данных.

Основные проблемы при классификации данных – это определение критериев отнесения входных данных к конкретному классу и построение классификатора, способного наиболее точно разделить входные данные на различные классы (кластеры).

Классификаторы обычно тренируются на имеющемся массиве входных данных, после чего применяются для анализа новых входных данных. В результате анализа принимается решение о принадлежности новых данных к одному из известных классов (кластеров).

Основные показатели точности классификации – вероятность правильного определения принадлежности к заданному классу (Detection Rate – DR) и вероятность ложного определения принадлежности к заданному классу (False Alarm Rate – FAR). Для более детальной оценки качества классификации применяется матрица ошибок (confusion matrix).

Краткая характеристика набора данных KDD cup'99. Для проверки эффективности различных методов синтеза классификаторов при решении задач обнаружения сетевых атак организацией «ACM SIGKDD» в 1999 г. было проведено соревнование KDD cup'99 [1].

Набор данных KDD cup'99 [1] содержит записи из журналов событий реальной компьютерной сети. Каждая строка соответствует одному событию (сетевому соединению). В начале каждой строки идёт 41 значение (атрибуты), описывающее различные характеристики соединения, в конце каждое событие помечено как нормальное (normal) или как вредоносное (например, snmpgetattack – попытка использования протокола SNMP для вторжения). Все сетевые атаки разделены на 4 основных класса: Probe (сканирование), DoS (отказ в обслуживании), R2L (внедрение) и U2R (повышение привилегий). Каждый класс, в свою очередь, разделён на подклассы (в зависимости от конкретной реализации сетевой атаки).

Для тестирования классификаторов данные разделены на две части: «kddcup.data_10_percent» и «Corrected KDD». Первая часть данных необходима для тренировки классификатора, вторая для проверки его эффективности. Для работы с исследуемым нейросетевым классификатором каждое 41 входное значение преобразовано в 8-битовую форму. В результате каждое событие представлено в виде вектора длиной 328 бит.

Предлагаемая модель искусственного нейрона. Для улучшения характеристик распознавания и уменьшения используемых вычислительных ресурсов в качестве элемента классификации предлагается использовать новую модель искусственного нейрона (CyberNeuron) [2]. В отличие от классических моделей искусственных нейронов данный тип нейрона использует табличные подстановки вместо операции умножения входных значений на весовые коэффициенты. Это позволило значительно увеличить информационную ёмкость отдельного нейрона (элемента нейросети), что в результате даёт ряд существенных преимуществ перед классической моделью.

Пример работы кибернейрона [2] приведен на рис. 1. Модель состоит из двух блоков: блока табличной подстановки и блока суммирования.

Каждое входное значение подаётся на соответствующую таблицу подстановки (sbox). Дискретное значение входа интерпретируется как индекс ячейки таблицы (index), а выходом является значение, хранящееся в ячейке. Выходы со всех таблиц подстановок суммируются, в результате формируется выход нейрона.

$$output = \sum_{i=1}^{N} sbox_i[input_i]$$

Модель имеет следующие особенности:

- таблица подстановки соответствует синапсу биологического нейрона, а операция суммирования – суммарному воздействию синапсов на возбуждение нейрона. В соответствии с данной моделью каждый синапс, в зависимости от входного воздействия, может быть как тормозящим, так и возбуждающим (каждая ячейка таблицы может хранить как отрицательные, так и положительные числа);

- размерность таблицы подстановки (количество ячеек таблицы) соответствует диапазону входных значений;
- размерность хранящихся в ячейках таблицы чисел может быть произвольной и зависит от решаемой задачи;
- в данной модели отсутствует выходная нелинейная функция (функция активации). Это объясняется следующим: таблицей подстановки можно описать любую дискретную функцию, в том числе и функцию умножения дискретных чисел, и дискретные функции активации. Поэтому при помощи данной модели можно также реализовать формальный нейрон, использующий дискретные вычисления.

Следует отметить, что в качестве операции объединения вместо суммирования можно применять и другие методы объединения, например, подавать полученные значения с таблиц подстановок на входы таблиц подстановок искусственных нейронов последующих слоёв нейросети.

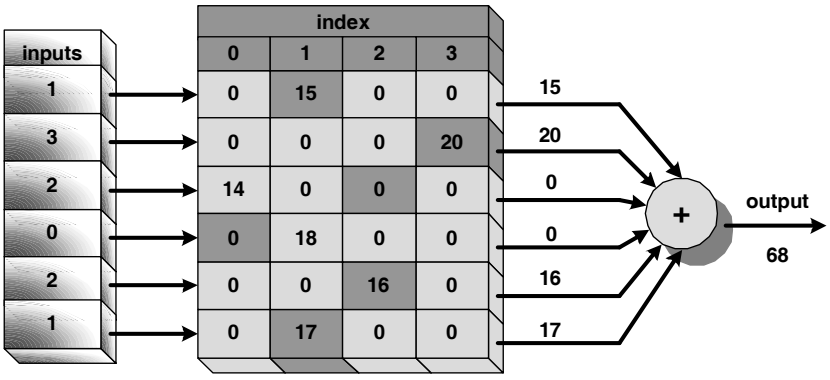

Рис. 1. Пример работы кибернейрона, обученного на образ (1,3,0,1,2,1), при подаче на его вход образа (1,3,2,0,2,1)

Главные преимущества новой модели искусственного нейрона перед формальным нейроном:

1. Обладает бо́льшей информационной ёмкостью, которая зависит от размерности используемых входных таблиц подстановок и может достигать значения миллионов «образов» при использовании на обычных персональных компьютерах (формальный нейрон – обычно не более нескольких «образов»).

2. Гораздо меньший размер искусственной нейросети при одинаковом качестве распознавания «образов», вследствие гораздо большей информационной ёмкости отдельного нейрона.

3. Эффективная и простая реализация в программном и аппаратном видах.

4. Намного меньшее количество итераций (эпох) обучения при сопоставимом объёме обучающей выборки и значительно выше скорость работы нейросети.

Основные недостатки новой модели:

- требуют разработки принципиально других методов тренировки нейросети [2];
- неприменимы такие широко распространённые методы обучения, как метод обратного распространения ошибки.

Применение кибернейрона для классификации данных KDD cup'99. На основе предложенной модели искусственного нейрона был разработан метод синтеза кибернейронного «дерева» (*Cyberneuron Tree*). Данный метод заключается в поэтапном наращивании нейросети. На каждом этапе к нейросети добавляется одна таблица подстановки, для которой определяются параметры соединений и содержимое.

Параметры соединений определяются эволюционным методом, для чего каждый этап разбивается на итерации. В течение каждой итерации определяется наилучшее текущее соединение для одного входа таблицы подстановки. Для этого отдельно выбранный вход таблицы подстановки поочерёдно соединяется со всеми доступными входными битами с одновременным обновлением содержимого таблицы. Выбирается то соединение, которое даёт наилучший прирост для параметра PROP. Параметр PROP отражает долю правильного распознавания входных данных классификатором и задаётся в следующем виде:

$$PROP = \frac{Ngood}{Ngood + Nbad} = \frac{k \cdot K11 + K00}{k \cdot K11 + k \cdot K10 + K01 + K00},$$

где

$Ngood = k \cdot K11 + K00$ – количество правильных срабатываний классификатора;

$Nbad = k \cdot K10 + K01$ – количество неправильных срабатываний классификатора;

K11 – количество правильных срабатываний для «1» (DR);

K10 – количество неправильных срабатываний для «1»;

K01 – количество неправильных срабатываний для «0» (FAR);

K00 – количество правильных срабатываний для «0»;

k = N0 / N1 – выравнивающий коэффициент;

N0 – количество событий вида «0»;

N1 – количество событий вида «1»;

«0» – входное событие не соответствует заданному классу;

«1» – входное событие соответствует заданному классу.

Таким способом поочерёдно определяются параметры соединения для каждого входа таблицы подстановки.

В табл. 1 приведено сравнение эффективности синтезируемых нейросетей (при различной размерности таблиц подстановки) с известными результатами.

Таблица 1

Классификатор	Norm	Probe	DoS	U2R	R2L	DR/ FAR
KDD 99 winner [3]	99.5	83.3	97.1	13.2	8.4	DR
	27.0	35.2	0.1	28,6	1.2	FAR
PNrule [4]	99.5	73.2	96.9	6.6	10.7	DR
	27.0	7.5	0.05	89.5	12.0	FAR
Multi-class SVM [5]	99.6	75	96.8	5.3	4.2	DR
	27.8	11.7	0.1	47.8	35.4	FAR
Layered Conditional Random Fields [6]	-	98.60	97.40	86.30	29.60	DR
	-	0.91	0.07	0.05	0.35	FAR
Columbia Model [7]	-	96.7	24.3	81.8	5.9	DR
Decision Tree [8]	-	81.4	60.0	58.8	24.2	DR
BSPNN [9]	99.8	99.3	98.1	89.7	48.2	DR
	3.6	1.1	0.06	0.03	0.19	FAR

Окончание табл. 1

Классификатор	Norm	Probe	DoS	U2R	R2L	DR/ FAR
Cyberneuron Tree, 4bit	-	83.7	97.13	94.73	33.51	DR
		0.49	0.22	0.72	0.22	FAR
Cyberneuron Tree, 6bit	-	80.78	97.16	56.14	24.05	DR
		0.40	0.32	0.27	0.04	FAR
Cyberneuron Tree, 8bit	-	65.28	97.28	-	-	DR
		1.7	0.103	-	-	FAR

Из приведенных данных видно, что синтезированные классификаторы в основном показывают сравнительные и более лучшие результаты (например, для атак вида R2L).

Выводы. Предлагаемый метод синтеза классификатора на основе новой модели искусственного нейрона позволяет формировать классификаторы, имеющие характеристики на уровне наилучших современных решений. При этом получаемые классификаторы являются вычислительно-эффективными – для их работы требуется небольшое количество операций табличной подстановки и сложений, а также небольшой объём оперативной памяти.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Официальный сайт организации «ACM Special Interest Group on Knoweledge Discovery and Data Mining». www.kdd.org.
2. *Поликарпов С.В., Дергачёв В.С., Румянцев К.Е., Голубчиков Д.М.* Новая модель искусственного нейрона: кибернейрон и области его применения. http://arxiv.org/abs/0907.0229.
3. *Pfahringer B.* Winning the KDD99 Classification Cup: Bagged Boosting // SIGKDD Explorations. – 2000. – Vol. 1. – P. 65-66.
4. *Agarwal R. and Joshi M.V.* PNrule: A New Framework for Learning Classifier Models in Data Mining // in A Case-Study in Network Intrusion Detection, 2000.
5. *Ambwani T.* Multi class support vector machine implementation to intrusion detection // in Proc. of IJCNN. – 2003. – P. 2300-2305.
6. *Gupta K.K., Nath B., Kotagiri R.* Layered Approach using Conditional Random Fields for Intrusion Detection // IEEE Transactions on Dependable and Secure Computing. – 2008. – Vol. 5.
7. *Lee W., Stolfo S.* A Framework for Constructing Features and Models for Intrusion Detection Systems // Information and System Security. – 2000. – Vol. 4. – P. 227-261.
8. *Lee J.H., Sohn S.G., Ryu J.H., Chung T.M.* Effective Value of Decision Tree with KDD 99 Intrusion Detection Datasets for Intrusion Detection System // in 10th International Conference on Advanced Communication Technology. – 2008. – Vol. 2. – P. 1170-1175.
9. *Tich Phuoc Tran, Longbing Cao, Dat Tran, Cuong Duc Nguyen.* Novel Intrusion Detection using Probabilistic Neural Network and Adaptive Boosting // International Journal of Computer Science and Information Security. – 2009. – Vol. 6, № 1. – P. 83-91.

Статью рекомендовал к опубликованию д.т.н., профессор Д.А. Безуглов.

Поликарпов Сергей Витальевич – Технологический институт федерального государственного автономного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге; e-mail: polikarpovsv@gmail.com; 347928, г. Таганрог, ул. Чехова, 2; тел.: 88634315507; кафедра информационной безопасности телекоммуникационных систем; к.т.н.; доцент.

Polikarpov Sergej Vital'evich – Taganrog Institute of Technology – Federal State-Owned Autonomy Educational Establishment of Higher Vocational Education "Southern Federal University"; e-mail: polikarpovsv@gmail.com; 2, Chekhov street, Taganrog, 347928, Russia; phone: +78634315507; the department of information security of telecommunication systems; cand. of eng. sc.; associate professor.

УДК 621.391.25(075)

И.Л. Трунов, У.Д. Линенко, А.В. Пустоварова

ИСПОЛЬЗОВАНИЕ СВОЙСТВ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ФИБОНАЧЧИ В СИСТЕМАХ ПОМЕХОУСТОЙЧИВОГО КОДИРОВАНИЯ

Рассмотрены особенности помехоустойчивых кодов, основанных на последовательностях Фибоначчи; обосновано применение максимальной и минимальной форм кодировки сообщений; предложены варианты структур помехоустойчивых систем передачи информации, использующих эти коды. Используя свойства последовательностей Фибоначчи в приемной части системы передачи информаци, возможно обнаружить ошибку уже по двум последовательно принятым разрядам кода, причем локализовать ее с точностью до двух символов. Это позволяет сделать вывод о достоверности принятого сообщения по его части, не дожидаясь конца сообщения.

Помехоустойчивое кодирование; последовательности Фибоначчи.

I.L. Trunov, U.D. Linenko, A.V. Pustovarova

USAGE OF SEQUENCES FIBONACCI PROPERTIES IN NOISE-RESISTANT CODING SYSTEMS

Noiseproof codes features based on sequences of Fibonacci are considered; application of the maximum and minimum forms of messages coding is justified; structures options of the noiseproof information transmission systems using these codes are offered. Using properties of sequences of Fibonacci in a receiving part of information transmission system, it is possible to find an error already on two sequentially accepted code discharges, and to localize it to within two characters. It allows to draw an output on the accepted message reliability by its part, without waiting the message end.

Noise-resistant coding; sequences Fibonacci.

В настоящее время темпы развития телекоммуникационных систем стали предпосылкой для появления принципиально новых способов кодирования сообщений. Несмотря на рост мощности вычислительной техники, актуальным остается вопрос построения простых алгоритмов коррекции ошибок.

Помехоустойчивое кодирование передаваемой информации позволяет в приемной части системы обнаруживать и исправлять ошибки.

Под р-кодом Фибоначчи понимается следующий способ представления натурального числа N, показанный в формуле (1):

$$N = anFp(n) + an-1Fp(n-1) + ... + aiFp(i) + ...)a1Fp, \quad (1)$$

где ai = {0, 1} – двоичная цифра i-го разряда представления; n – разрядность представления; Fp(i) – р-число Фибоначчи, задаваемое с помощью следующих рекуррентных формул (2) и (3):

$$Fp(i) = Fp(i-1) + Fp(i-p-1), \quad (2)$$

$$Fp(1) = Fp(2) = ... = Fp(p+1) = 1, \quad (3)$$

где p – целое неотрицательное число, принимающее значение из множества {0, 1, 2, 3 ...} [1, 2].