УДК 004.056.5

РАЗРАБОТКА ПРОЦЕДУРЫ ОПРЕДЕЛЕНИЯ ОЖИДАЕМОГО ВЫХОДНОГО СИГНАЛА НЕЙРОСЕТЕВОЙ МОДЕЛИ РАСПОЗНАВАНИЯ КИБЕРАТАК

Айтчанов Б.Х., Бапиев И.М.

Казахский национальный исследовательский технический университет им. К.И. Сатпаева,
Алматы, e-mail: bekait@rambler.ru

В данной статье показана процедура разработки определения ожидаемого выходного сигнала нейросетевой модели, предназначенной для распознавания атак в киберпространстве на сетевые ресурсы информационных систем. Особенностью разработанной процедуры является учет в ожидаемом выходном сигнале близости распознаваемых эталонов кибератак и эталонов распознаваемых безопасных состояний. Учитывать близость эталонов предлагается с помощью экспертного оценивания. Для верификации полученных теоретических результатов проведены экспериментальные исследования, в ходе которых была построена и обучена нейросетевая модель, предназначенная для распознавания двух видов кибератак и одного безопасного состояния. Основная гипотеза эксперимента – использование разработанной процедуры позволяет уменьшить количество учебных итераций, необходимых для достижения заданной ошибки обучения. Проведенные численные эксперименты показали, что применение разработанной процедуры обучения нейросетевой модели, предназначенной для распознавания сетевых кибератак, позволяет примерно на 20% уменьшить количество учебных итераций, что в свою очередь позволяет повысить оперативность создания нейросетевой модели.

Ключевые слова: нейронная сеть, обучающая выборка, нейросетевая модель, распознавание кибератак

DEVELOPMENT OF PROCEDURES FOR DETERMINING THE EXPECTED OUTPUT NEURAL NETWORK MODEL RECOGNITION CYBERATTACKS

Aitchanov B.H., Bapiev I.M.

Kazakh National Research Technical University after K.I. Satpaev, Almaty, e-mail: bekait@rambler.ru

This article shows the procedure for developing the definition of the expected output signal of a neural network model designed to recognize attacks in cyberspace on network resources of information systems. The peculiarity of the developed procedure is the registration in the expected output signal of proximity of the recognized standards of cyber attacks and standards of recognized safe states. To take into account the proximity of the standards is proposed with the help of expert evaluation. To verify the theoretical results obtained, experimental studies were carried out, during which a neural network model designed to recognize two types of cyber attacks and one safe state was constructed and trained. The main hypothesis of the experiment – using the developed procedure allows to reduce the number of training iterations necessary to achieve a given learning error. Numerical experiments have shown that application of the developed procedure for learning the neural network model designed for recognition of network cyberattacks allows to reduce by approximately 20% the number of training iterations, which in turn allows increasing the efficiency of creating a neural network model.

Keywords: neural network, learning sample, the neural network model, the recognition of cyber attacks

На протяжении последних нескольких лет важным направлением повышения уровня защищенности сетевых ресурсов информационных систем (РИС) является разработка и внедрение эффективных нейросетевых средств (НСР) распознавания кибератак на эти ресурсы [1 – 3, 6]. Хотя анализ источников [3, 6 – 8] и указывает на достаточно большой научно-практический задел в этом направлении, однако этот же анализ указывает на недостаточную эффективность обучения нейросетевых моделей (НСМ), являющихся основой указанных НСР. Из-за этого недостатка повышается время построения НСР и уменьшается точность распознавания кибератак. Также результаты [3, 4] позволяют утверждать, что повысить эффективность обучения современных НСМ возможно путем повышения качества учебных примеров за счет отображения в ожидаемом выходном сигнале близости эталонов

распознаваемых состояний защищенности. Таким образом, целью данного исследования является разработка формализованной процедуры определения ожидаемого выходного сигнала для нейросетевой модели распознавания кибератак на сетевые ресурсы информационных систем.

Поскольку предусмотрено распознавать кибератаки на основании анализа соответствующих характеристик РИС, то предлагается, чтобы близость этих характеристик отображалась в мере схожести кибератак между собой. Также, базируясь на результатах [2, 3], определено, что в базовом случае выходной сигнал НСМ реализуется с помощью одного нейрона в выходном слое $N_y=1$, где N_y – количество нейронов в выходном слое.

При использовании сигмоидальной функции активации нейрона выходного слоя, что характерно для НСМ на базе мно-

гослоино персептрона, выходной сигнал у находится в пределах от 0 до 1.

При распознавании каждой из возможных кибератак, а также каждому из возможных безопасных состояний сетевого РИС ставится в соответствие некоторый диапазон величин выходного сигнала. В дальнейшем для краткости распознаваемые виды кибератак и распознаваемые безопасные состояния РИС будем называть состояниями защищенности.

Не теряя общности рассуждений, можно предположить, что величины диапазонов для разных состояний защищенности разные. Кроме того, для учебных примеров, которые отвечают эталонам состояний защищенности, выходной сигнал будет равен середине указанного диапазона.

В случае использования сигмоидальной функции активации нейронов выходного слоя и при условии равномерного квантования диапазона возможных значений у ожидаемый выходной сигнал для эталона произвольного i -го состояния защищенности рассчитывается так:

$$y_{s_i} = \frac{1}{K_s} i - \frac{0,5}{K_s} = \frac{i-0,5}{K_s}, \quad (1)$$

где K_s – количество распознаваемых состояний защищенности, i – номер состояния защищенности.

Схожесть состояний защищенности в выражении (1) возможно учесть только за счет того, что схожие состояния защищенности должны иметь близкие номера. Для этого необходимо провести числовую оценку близости состояний защищенности. Известные аналитические методы такого расчета [4] отличаются большой сложностью и низкой надежностью, что затрудняет их эффективное использование для распознавания кибератак на сетевые РИС. В то же время, анализ и распознавание кибератак на сетевые РИС – это задачи, которые достаточно эффективно решаются экспертами в области защиты информации [3, 5]. Поэтому представляется целесообразным определять числовую оценку степени схожести параметров кибератак и параметров безопасных состояний на основе экспертных данных.

Базируясь на результатах [5], предлагается использовать статистические методы обработки экспертных данных. В этом случае полученные от экспертов количественные данные обрабатываются с целью оценки коллективного мнения экспертной группы, оценки согласованности мнений экспертов и оценки их компетентности. Для определения оценок используются статистические

методы точечного и интервального оценивания. Для этого рекомендуется, чтобы количество экспертов было не менее 10.

Рассмотрим процесс экспертного оценивания степени близости состояний защищенности. Пусть в результате опроса экспертной группы, которая состоит из m участников, получены следующие данные:

$$\begin{matrix} x_{1,1}, & \dots & x_{n,1} & \dots & x_{N,1} \\ \dots & \dots & \dots & \dots & \dots \\ x_{1,m} & \dots & x_{n,m} & \dots & x_{N,m} \\ \dots & \dots & \dots & \dots & \dots \\ x_{1,M} & \dots & x_{n,M} & \dots & x_{N,M} \end{matrix}, \quad (2)$$

где $x_{n,m}$ – оценка степени схожести n -го объекта (состояния защищенности) m -м экспертом, N – количество объектов (состояний защищенности), M – количество экспертов.

Средняя коллективная оценка n -го состояния защищенности рассчитывается с помощью формулы:

$$x_n = \frac{1}{M} \sum_{m=1}^M x_{n,m}, \quad (3)$$

где $x_{n,m}$ – оценка степени схожести n -го состояния защищенности m -м экспертом, $n=1 \dots N$.

Дисперсия средней коллективной оценки определяется так:

$$\sigma^2 = \frac{1}{M-1} \sum_{m=1}^M (x_{n,m} - x_n)^2. \quad (4)$$

Для определения статистической значимости полученных результатов необходимо указать доверительный интервал, в который оцениваемая величина попадает с заданной доверительной вероятностью P .

Задав вероятность ошибки P_n (с учетом уровня значимости), можно определить интервал, в который оцениваемая величина попадает с вероятностью $(1 - P_n)$:

$$I_{x_n} = (x_n - \varepsilon_{pn}, x_n + \varepsilon_{pn}). \quad (5)$$

Величина ε_{pn} определяет границы доверительного интервала и рассчитывается так:

$$\varepsilon_{pn} = t_p \frac{\sigma_n}{\sqrt{M}}, \quad (6)$$

где t_p – коэффициент, зависящий от заданной доверительной вероятности P .

Считается, что оцениваемая величина имеет нормальное распределение с центром x_i и дисперсией σ . Коэффициент t_p имеет распределение Стьюдента с $(N-1)$ степеня-

ми свободы и определяется с помощью табличных значений [5].

Степень согласованности экспертных мнений определяется с помощью коэффициента вариации γ_n , который рассчитывается по формуле:

$$\gamma_n = \frac{\sigma_n}{x_n}. \quad (7)$$

Рассчитанный с помощью выражения (7) коэффициент вариации γ_n определяет относительную величину диапазона изменения оценок экспертов относительно среднего значения коллективной оценки x_n . При полной согласованности мнений экспертов, когда все $x_{n,m} = x_n$, коэффициент вариации $\gamma_n = 0$.

Отметим, что полученные выражения (1)–(7) составляют основу математического обеспечения отдельных операций процедуры определения ожидаемого выходного сигнала нейросетевой модели распознавания кибератак на сетевые ресурсы. При этом порядок выполнения этих операций соответствует номерам указанных выражений.

Для верификации полученных теоретических результатов проведены экспериментальные исследования, в ходе которых была построена и обучена HCM, предназначенная для распознавания двух видов кибератак и одного безопасного состояния. Основная гипотеза эксперимента – использование разработанной процедуры позволяет уменьшить количество учебных итераций, необходимых для достижения заданной ошибки обучения.

В качестве источника данных для HCM использована база данных KDD-99, в которой записаны значения 41 параметра для сетевых соединений, соответствующих 22 видам кибератак и одному безопасному состоянию. В эксперименте распознавались четыре типа кибератак вида R2L, которые направлены то, чтобы незарегистрированный пользователь получил доступ к компьютеру со стороны удаленной машины. Типы распознаваемых кибератак: buffer_overflow, perl, loadmodule и rootkit. Также предусмотрено распознавание безопасного соединения.

Построение HCM реализовано на базе методологии, разработанной в [1, 4, 5]. В качестве базового вида HCM использован двухслойный персептрон с $N_x = 41$ входным и $N_y = 1$ выходным нейроном. Количество входных параметров выбрано, исходя из структуры записей KDD-99, а количество выходных параметров мотивировано упрощением структуры модели. Выбор количе-

ства учебных примеров $P=1000$ базировался на использовании выражения (16), обоснованного в [6]:

$$P_{\min} > 20_{x_s}, \quad (16)$$

где $P_{\min}$ – минимальное количество учебных примеров.

При формировании учебной выборки предусмотрено одинаковое количество примеров для каждого из распознаваемых состояний защищенности.

Количество скрытых нейронов $N_s = 405$ рассчитано с использованием выражения (17), также определенного в [1, 6]:

$$N_s = \text{Round} \left(\frac{2\sqrt{PN_x}}{N_y} \right), \quad (17)$$

где $\text{Round}(X)$ – операция определения ближайшего целого числа от аргумента X .

Проведено две серии численных экспериментов, направленных на определение количества учебных итераций HCM для достижения безошибочного распознавания учебных примеров. В первом эксперименте ожидаемый выходной сигнал определялся с помощью выражения (7) с использованием предпосылки, что состояния защищенности ранжированы по алфавиту. Во втором эксперименте ожидаемый выходной сигнал определялся с помощью предложенной процедуры.

В результате экспериментов установлено, что при использовании предложенной процедуры определения ожидаемого значения выходного параметра количество учебных итераций для достижения безошибочного запоминания HCM всех учебных примеров уменьшилось приблизительно на 20%, что подтверждает принятую гипотезу. Также в первом приближении можно считать, что за счет уменьшения количества учебных итераций примерно на 20% возрастает оперативность создания HCM.

Таким образом, в результате проведенных исследований разработана процедура определения ожидаемого выходного сигнала нейросетевой модели распознавания кибератак на сетевые ресурсы, которая за счет учета в таком сигнале близости распознаваемых эталонов кибератак и эталонов распознаваемых безопасных состояний позволяет повысить оперативность создания указанных моделей. При этом проведенные численные эксперименты показали, что применение разработанной процедуры для обучения нейросетевой модели, позволяет приблизительно на 20% повысить оперативность создания такой модели.

Список литературы

1. Абрамов Е.С. Разработка и исследование методов построения систем обнаружения атак: дис. ... канд. техн. наук : 05.13.19. – Таганрог, 2005. – 199 с.
2. Артеменко А.В. Анализ нейросетевых методов распознавания компьютерных вирусов / А.В. Артеменко, В.А. Головкин // Материалы секционных заседаний. Молодежный инновационный форум «ИНТРИ» – 2010. – Минск: ГУ «БелИСА», 2010. – 239 с.
3. Бапиев И.М. Применение нейронной сети с радиальными базисными функциями для распознавания скриптовых вирусов / И.М. Бапиев, Б.С. Ахметов, А.Г. Корченко, И.А. Терейковский // II международная научно-практическая конференция «Актуальные вопросы обеспечения кибербезопасности и защиты информации». – Киев: 24–27 февраль 2016. – С. 20–23.
4. Бапиев И.М. Разработка критериев оценки эффективности нейросетевых средств распознавания кибератак на сетевые ресурсы информационных систем / И.М. Бапиев, А.Г. Корченко, И.А. Терейковский // Тезисы докл. IV Международной научно-практической конференции «Global and regional problems of informatization in society and nature using», – Киев: 23–24 июнь 2016, С. 80–82.
5. Большев А.К. Алгоритмы преобразования и классификации трафика для обнаружения вторжений в компьютерные сети: Автореф. дисс. ... канд. техн. наук: спец. 05.13.19 – Методы и системы защиты информации, информационная безопасность. – СПб., 2011. – 36 с.
6. Гамаюнов Д.Ю. Обнаружение компьютерных атак на основе анализа поведения сетевых объектов: Автореф. дисс. ... канд. техн. наук: спец. 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей. – М., 2007. – 11 с.
7. Емельянова Ю.Г. Нейросетевая технология обнаружения сетевых атак на информационные ресурсы / Ю.Г. Емельянова, А.А. Талалаев, И.П. Тищенко, В.П. Фраденко // Программные системы: теория и приложения. – 2011. – №3(7). – С. 3–15.
8. Корченко А. Нейросетевые модели, методы и средства оценки параметров безопасности Интернет-ориентированных информационных систем: монография / А. Корченко, И. Терейковский, Н. Карпинский, С. Тынымбаев. – К.: ТОВ «Наш Формат». 2016. – 275 с.