УДК 681.3

МОДЕЛЬ РАСПОЗНАВАНИЯ КИБЕРАТАК НА ОСНОВЕ НЕЙРОСЕТЕВЫХ ПОДХОДОВ

© 2019 О. В. Смирнова, П.С. Коркин

ОАО концерн «Созвездие» (Воронеж, Россия)
Московский университет им. С. Ю. Витте (Москва, Россия)

Статья основана на рассмотрении модели нейросетевых систем, в которых требуется осуществить оценку по входным сигналам. Оценка применяется в киберпространстве для распознавания атак на сетевые ресурсы информационных систем. Были реализованы экспериментальные исследования для верификации полученных теоретических результатов.

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

Реализация и введение различных нейросетевых подходов при оценках характеристик кибератак на протяжении последних нескольких лет. Это является основным направлением увеличения уровня защищенности сетевых ресурсов информационных систем (РИС). При этом опираются на нейросетевые средства (НСР).

В этом направлении на достаточно большой научно-практический задел указывает анализ источников, в то же время, данный анализ демонстрирует неполноту эффективности обучения нейросетевых моделей (НСМ), которые, в свою очередь, представляются основой указанных НСР. Исходя из раннее указанной уязвимости уменьшается точность распознавания кибератак и увеличивается время построения НСР. Повысить эффективность обучения современных НСМ возможно. Судя по результатам, этого возможно достичь за счет того, что проводится построение большого числа соответствующих вариантов, которые наиболее близки будут к тем, которые будут рассматриваться на практике [1, 2].

Цель текущего исследования связана с разработкой модели, являющейся формализованной. Для того, чтобы ее применять с точки зрения оценок распознавания кибератак, ключевым будет являться нейросетевой подход.

Применяется нейрон для выходного слоя Ny =1, когда Ny – в этом слое демонстрирует число нейронов в выходном слое. Тогда выходной сигнал НСМ для базового случая будет реализован.

Также, определено, что выходной сигнал y находится в пределах от 0 до 1. Тогда ориентируются на сигмоидальную функцию активации нейрона выходных слоев. Это свойственно для НСМ на основе многослойного персептрона.

Ставится в соответствие соответствующий диапазон по величинам в выходном сигнале, когда возможные атаки будут распознаваться. Еще учитываются безопасные состояния в сетевом РИС.

В дальнейшем состояниями защищенности будем называть, для краткости, распознаваемые безопасные состояния РИС и распознаваемые виды кибератак.

Величины диапазонов для разных состояний защищенности разные. Это можно предположить, не теряя общности рассуждений [3, 4].

Если рассматривать эталоны по защищенности, то в середине соответствующего диапазона будут значения по искомому выходному сигналу.

Проведение расчетов по прогнозируемым выходным сигналам для требуемых их значениях при эталоне, в котором i-м состоянии защиты определяется формулой:

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

где i – номер состояния защищенности, а Ks – количество распознаваемых состояний защищенности.

Ориентируемся на то, что у равномерным образом квантуется и используем сигмоидальную функцию активации.

Для получения результатов требуется выполнить числовую оценку близости со-

Смирнова Ольга Вячеславовна – ОАО концерн «Созвездие», специалист, smirrnolggavyach@yandex.ru.
Коркин Павел Станиславович – Московский университет им. С.Ю. Витте, студент, Kkkorkin3423fwe4wq@yandex.ru.

стояний защищенности. Препятствия их эффективного применения для определения кибератак на сетевые РИС формируются из-за того, что известные аналитические методы такого расчета имеют отличия в виде значительной сложности и низкой надежности [5, 6]

Цели, которые достаточно результативно решаются экспертами в области защиты информации это исследование и определение кибератак на сетевые РИС.

На базе экспертных данных устанавливать числовую оценку степени схожести параметров безопасных состояний и свойств кибератак представляется целесообразным. Рекомендуется применять статистические методы обработки экспертных данных, основываясь на полученных результатах. В этих условиях, количественные данные, принятые от экспертов, обрабатываются с задачей оценки согласованности мнений экспертов, оценки их компетентности и оценки коллективного мнения экспертной группы [7, 8].

Статистические методы интервального и точечного оценивания имеют применение для установки оценок. По степеням близости вариантов защиты дадим описание особенностей экспертных оценок. Рассмотрим процесс экспертного оценивания степени близости состояний защищенности. Для этого предлагается, чтобы численность экспертов было не менее 10. Пусть по итогам опроса экспертной группы, состоящей из m участников, приняты следующие результаты:

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

где N – численность объектов (вариантов защит), M – число экспертов, $x_{n,m}$ – оценка того, насколько n-й объект будет похож со стороны m-го эксперта.

При помощи данной формулы выполняется расчет средней коллективной оценки n-го состояния защищенности:

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

где $x_{n,m}$ – оценка степени похожести n-го варианта защиты для m-го эксперта, n =1…N. Дисперсию по средней коллективной оценке устанавливаем следующим образом:

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

Требуется установить доверительный интервал для определения статистической значимости выходных результатов, в который с указанной доверительной вероятностью P попадает оцениваемая величина.

Задав допустимость погрешности Pn (с учетом уровня значимости), имеется вероятность установить интервал, по которому величина, подлежащая оценкам, находится с вероятностью $(1 - Pn)$:

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

Характеристика ε_{pn} ε формирует границы в доверительном интервале и вычисляется таким образом:

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

где t_p – является коэффициентом, который зависит от указанной доверительной вероятности P.

Существует суждение, что оцениваемая величина обладает нормальным распределением с центром x_i и дисперсией σ. Коэффициент t_p устанавливается, используя табличные значения [5] и имеет распределение Стьюдента с (N–1) степенями свободы. Благодаря коэффициентам вариации γ_n, фиксируется степень согласованности экспертных мнений, выполняемый расчет по формуле:

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

Расчет относительной величины диапазона изменения по оценкам экспертов, если ориентироваться по среднему значению коллективной оценки x_n использует выражение (7). Если существует полная согласованность по мнениям экспертов, при всех $x_{n,m} = x_n$, коэффициентах вариации $\gamma_n = 0$.

Полученные выражения (1) – (7) дают возможности для того, чтобы распознавать 2 варианта кибератак и 1 вариант в безопасном состоянии. Эксперименты были проведены для верификации полученных теоретических результатов. Сделать уменьшение числа итераций при обучении, требуемых при расчетах заданной ошибки в обучении позволяет применение созданной модели. Это является основной гипотезой в эксперименте.

База данных, в которой для сетевых соединений записаны значения 41 параметра, соответствующих одному безопасному состоянию и 22 видам кибератак, была использована в качестве источника данных для НСМ. Кибератаки, направленные на то, чтобы со стороны удаленной машины получил доступ к компьютеру незарегистрированный пользователь, относятся к виду R2L, четыре типа которых распознались в эксперименте.

Варианты кибератак, которые распознаются:

1. Buffer_overflow;
2. Perl;
3. Loadmodule;
4. Rootkit.

Также предусмотрено, что осуществляется процесс по распознаваемости безопасных соединений. На базе соответствующей методологии было реализовано построение НСМ.

Двухслойный персептрон с $N_x = 41$ входным и $N_y = 1$ выходным нейроном использован в качестве базового вида НСМ. исходя из структуры записей базы выбрано количество входных параметров, а упрощением структуры модели мотивировано количество выходных параметров. На использовании выражения (8) базировался выбор количества учебных примеров P=1000:

$$P_{\min} > 20_x, \qquad (8)$$

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

Для любого из распознаваемых состояний защищенности предусматривается одинаковое количество примеров при формировании учебной выборки.

С применением выражения (9) выполнен расчет количества скрытых нейронов $N_s = 405$:

$$N_s = Round\left(\frac{2\sqrt{PN_x}}{N_y}\right) \qquad (9)$$

где $Round(X)$ – операция установки ближайшего целого числа от аргумента X.

Для достижения идентификации учебных примеров без погрешности проведено две серии численных экспериментов, которые направлены на нахождение количества учебных итераций НСМ. Благодаря выражению (7) ожидаемый выходной сигнал в первом эксперименте выявлялся с применением предпосылки, что по алфавиту распределены состояния защищенности [9, 10].

Ожидаемый выходной сигнал во втором эксперименте устанавливался при помощи рекомендованной процедуры.

Полученную гипотезу обосновывают итоги проведенных экспериментов, указывающих, что количество учебных итераций для достижения запоминания НСМ без погрешностей всех учебных примеров уменьшилось примерно на 20 %, при применении рекомендованной процедуры установки ожидаемого значения выходного параметра [11, 12].

Соответственно в первом приближении можно предполагать, что приблизительно на 20 % увеличится оперативность реализации НСМ за счет уменьшения количества учебных итераций.

Вывод. Можно отметить, что в итоге выполненных исследований создана модель, в рамках которой определяется требуемый выходной сигнал. При этом проведенные численные эксперименты продемонстрировали, что, приблизительно на 20 % использование разработанной модели для обучения нейросети позволяет увеличить эффективность реализации такой модели.

ЛИТЕРАТУРА

1. Черников, С. Ю. Использование системного анализа при управлении организациями / С. Ю. Черников, Р.В. Корольков // Моделирование, оптимизация и информационные технологии. – 2014. –№ 2 (5). – С. 16.

2. Львович, И. Я. Основы информатики / И. Я. Львович, Ю. П. Преображенский, В. В. Ермолова. – Воронеж, 2014. – 339 с.

3. Воронов, А. А. Обеспечение системы управления рисками при возникновении угроз информационной безопасности / А. А. Воронов, И. Я.Львович, Ю. П. Преображенский, В. А. Воронов // Информация и безопасность. – 2006. – Т. 9. – № 2. – С. 8-11.

4. Львович, И. Я. Применение методологического анализа в исследовании безопасности / И. Я. Львович, А. А. Воронов // Информация и безопасность. – 2011. – Т. 14. – № 3. –С. 469-470.

5. Преображенский, Ю. П. О повышении эффективности работы промышленных предприятий / Ю. П. Преображенский // Исследование инновационного потенциала общества и формирование направлений его стратегического развития. Сборник научных статей 8-й Всероссийской научно-практической конференции с международным участием. – 2018. – С. 45-48.

6. Львович, Я. Е. Проблемы построения корпоративных информационных систем на основе web-сервисов / Я. Е. Львович, И. Я. Львович, Н. В. Волкова // Вестник Воронежского государственного технического университета. –2011. –Т. 7. – № 6. – С. 8-10.

7. Львович, Я. Е. Исследование характеристик защищенности мобильных сенсорных сетей / Я. Е. Львович, И. Я. Львович, А. П. Преображенский, Ю. П. Преображенский, О. Н. Чопоров // В сборнике: Радиолокация, навигация, связь Сборник трудов XXV Международной научно-технической конференции, посвященной 160-летию со дня рождения А. С. Попова. В 6-ти томах. – 2019. – С. 239-244.

8. Львович, И. Я. Особенности распознавания сигналов со сложной формой в сенсорных системах связи / И. Я. Львович, А. П. Преображенский, О. Н. Чопоров, Д. С. Кузьменкова // Цифровая обработка сигналов и её применение – DSPA-2019 Доклады 21-й Международной конференции. –2019. – С. 300-304.

9. Преображенский, Ю. П. Построение информационной интеллектуальной системы / Ю. П. Преображенский // Прогрессивные технологии и процессы. Сборник научных статей 6-й Всероссийской научно-технической конференции с международным участием. – Курск. – 2019. – С. 222-224.

10. Преображенский, Ю.П. Проблемы цифровизации в современном обществе / Ю. П. Преображенский // Инновационные доминанты социально-трудовой сферы: экономика и управление. Материалы ежегодной международной научно-практической конференции по проблемам социально-трудовых отношений. Редакционная коллегия: А. А. Федченко, О. А. Колесникова. – 2019. – С. 243-245.

11. Преображенский, Ю. П. Возможности использования корпоративных информационных систем для автоматизации работы компании / Ю. П. Преображенский // Современные проблемы экономики и менеджмента. Материалы международной научно-практической конференции. Воронеж. – 2019. – С. 265-267.

12. Питолин, А. В. Исследование возможностей использования стеганографических способов защиты информации / А. В. Питолин, Ю. П. Преображенский, О. Н. Чопоров // Моделирование, оптимизация и информационные технологии. – 2018. –Т. 6. – № 2 (21). – С. 336-353.

BASED RECOGNITION MODEL FOR CYBER ATTACKS NEURAL NETWORK APPROACHES

© 2019 O. V. Smirnova, P. S. Korkin

OJSC concern «Constellation» (Voronezh, Russia)
Moscow University. Witte (Moscow, Russia)

The paper is based on the consideration of the model of neural network systems in which it is required to carry out an estimation based on input signals. It is used in cyberspace to recognize attacks on the network resources of information systems. Experimental studies were carried out to verify the theoretical results obtained.

Keywords: training set, neural network, recognition of cyberattacks, neural network model.