www.volsu.ru

DOI: https://doi.org/10.15688/NBIT.jvolsu.2018.1.7

УДК 004.056.5
ББК 51.73

РАЗРАБОТКА ФОРМАЛЬНОЙ МОДЕЛИ ИССЛЕДОВАНИЯ ПРОГРАММ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА СОБЫТИЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Арина Валерьевна Никишова

Кандидат технических наук, доцент кафедры информационной безопасности,
Волгоградский государственный университет
arinanv@mail.ru, infsec@volsu.ru
просп. Университетский, 100, 400062 г. Волгоград, Российская Федерация

Светлана Владимировна Михальченко

Студент,
Волгоградский государственный университет
infsec@volsu.ru
просп. Университетский, 100, 400062 г. Волгоград, Российская Федерация

Аннотация. Рассмотрена проблема обеспечения информационной безопасности с точки зрения появления новых атак. Проанализированы программы для интеллектуального анализа событий информационной системы и сформулированы критерии для их оценки. Также разработана формальная модель исследования программ интеллектуального анализа событий информационной системы.

Ключевые слова: информационная безопасность, интеллектуальный анализ, STATISTICA Automated Neural Networks, Deductor Studio, Neural network toolbox, MemBrain Neural Network, NeuroSolutions.

По статистике, количество образцов новых атак, совершаемых в отношении информационных систем, растет (см. рисунок).

Неспособность обнаруживать новые образцы атак – недостаток современных систем обнаружения атак. Чтобы устранить данный недостаток внедряются интеллектуальные подходы к анализу данных для обнаружения атак [1].

Существует множество программ интеллектуального анализа данных, поэтому актуальным является выбор наилучшей программы.

В данной статье будет рассмотрено 5 программ, работающих с интеллектуальным анализом данных, а именно:

1. STATISTICA Automated Neural Networks (SANN) – программный пакет для создания и обучения нейронных сетей, который решает большой спектр задач [2].

2. Deductor Studio – это аналитическая платформа, которая позволяет на базе единой архитектуры пройти все этапы построения аналитической системы: от консолидации данных до построения моделей и визуализации полученных результатов [3].

3. Neural network toolbox (NNTool) – пакет расширения MATLAB, содержащий средства для проектирования, моделирования, разработки и визуализации нейронных сетей [2].

4. MemBrain Neural Network – представляет собой мощный графический редактор и симулятор нейронных сетей, поддерживающий нейронные сети различных архитектур любого размера.

5. NeuroSolutions – сверхсовременный программный пакет; совмещает модульный, с иконным представлением, интерфейс разработки нейронной сети, с реализацией усовершенствованных процедур обучения.

Чтобы оценить качество программ, предложены критерии для их оценки:

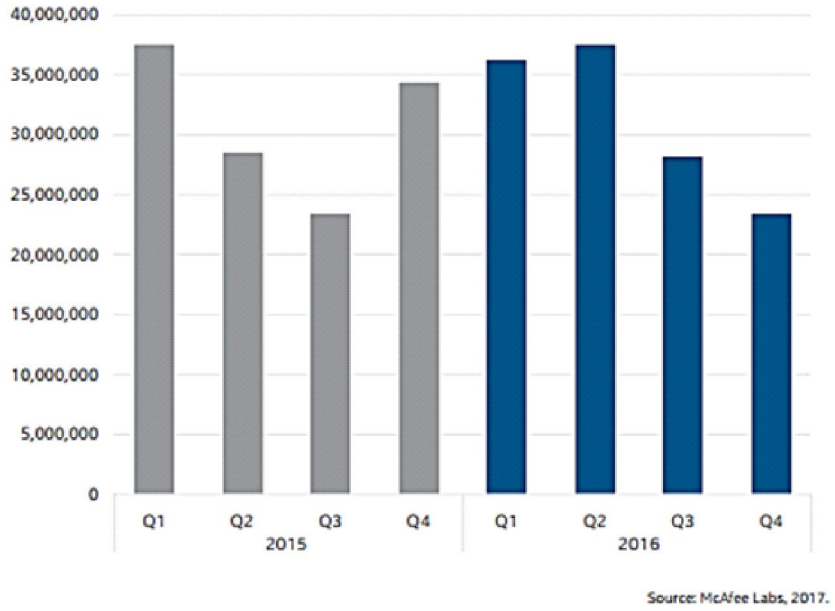
$К_1$. Скорость обучения нейронной сети – один из наиболее важных параметров, оценивающий эффективность программы. Он определяет величину изменения весовых коэффициентов связей между нейронами на каждом шаге обучения. Чтобы обеспечить наилучшую сходимость, шаг обучения нейронной сети должен стремиться к 0. Если выбирать шаг обучения очень маленький, то и время обучения возрастет. И наоборот: большой шаг займет хоть и небольшое количество времени, но нейронная сеть не будет сходиться. Поэтому некоторые разработчики внедряют в программу динамическую скорость обучения. Шаг зависит от определенных процессов и определяется системой.

$К_2$. Понятный графический интерфейс – нужно обеспечить удобный интерфейс пользователю, чтобы он смог легко определить, какие данные ему необходимы и куда их необходимо ввести. При этом интерфейс не должен содержать большое количество информации, так как это делает его более сложным для восприятия.

$К_3$. Наглядность информации – программа должна по окончании обучения и моделирования нейронной сети предлагать построить графики, диаграммы и другие варианты представления информации. Этот критерий тесно связан со вторым критерием, так как для пользователя важно представление конечной информации. От этого будет зависеть правильное восприятие информации и интерпретация результатов работы нейронной сети.

$К_4$. Реализация основных видов нейронных сетей и алгоритмов обучения – в процессе создания программы разработчик должен добавить возможность пользователю выбирать, каким видом нейронной сети он будет пользоваться и какой алгоритм обучения будет использован. Важной особенностью является реализация как можно большего числа стандартных видов нейронных сетей и алгоритмов для последующего обучения.

$К_5$. Создание своих структур нейронных сетей – должна предоставляться возможность создания своих нейронных сетей, позволяющая указывать такие параметры, как: тип сети,



Количество новых типов атак за 2015–2016 гг.

количество и размер скрытых слоев, алгоритм обучения нейронной сети и т.д.

$К_6$. Использование собственных алгоритмов – программа должна обладать возможностью подключения своих собственных алгоритмов обучения в виде программных модулей. Это функция делает программу более гибкой для разработчика.

$К_7$. Автоматическое формирование нейронной сети – такой критерий будет означать, что программа проводит анализ и подбирает наилучшие параметры автоматически, что облегчает использование такой программы.

$К_8$. Процедура импорта результатов – импорт полученных результатов должен предусматривать возможность сохранения результатов в различные файлы и приложения, что делает программу более удобной.

$К_9$. Генератор исходного кода – генератор кода может сгенерировать исходный системный программный код нейросетевых моделей на различных языках. Эта функция позволяет разработчику после создания и обучения сети генерировать код и встраивать его во внешние приложения.

$К_{10}$. Взаимосвязь (корреляция) событий – показывает, может ли программа строить и обучать нейронные сети, которые учитывают то, как события связаны между собой. Если в программе предусмотрена такая функция, то это позволяет обнаруживать атаки, которые состоят из нескольких шагов.

В таблице приведены значения критериев для выделенных программ.

Так как ни одна из программ не обладает наилучшим набором значений критериев, необходимо разработать подход к оценке программы интеллектуального анализа событий информационной системы для выбора из них наиболее рациональной.

Для этого нужно разработать математическую модель, реализующую эту оценку.

Сформируем вектор критериев

$$К = (К_1, К_2, К_3, К_4, К_5, К_6, К_7, К_8, К_9, К_{10}),$$

где $К_1$ – скорость обучения – принимает следующие значения:

$$К_1 = \begin{cases} 0, \text{низкая} \\ 0.5, \text{средняя} \\ 1, \text{высокая} \end{cases}$$

$К_2$ – понятный графический интерфейс – принимает следующие значения:

$$К_2 = \begin{cases} 0, \text{нет} \\ 1, \text{да} \end{cases}$$

$К_3$ – наглядность информации – принимает следующие значения:

$$К_3 = \begin{cases} 0, \text{низкая} \\ \frac{1}{2}, \text{средняя} \\ 1, \text{высокая} \end{cases}$$

$К_4$ – реализация основных видов нейронных сетей и алгоритмов обучения – принимает значения:

$$К_4 = \begin{cases} 0, \text{низкая} \\ \frac{1}{2}, \text{средняя} \\ 1, \text{высокая} \end{cases}$$

$К_5$ – создание своих структур нейронных сетей – принимает значения:

$$К_5 = \begin{cases} 0, \text{нет} \\ 1, \text{да} \end{cases}$$

Качественные значения критериев оценки программ интеллектуального анализа данных

Нейросетевые программы	Критерии оценки									
	$К_1$	$К_2$	$К_3$	$К_4$	$К_5$	$К_6$	$К_7$	$К_8$	$К_9$	$К_{10}$
STATISTICA Automated Neural Networks	Высокая	Да	Средняя	Высокая	Да	Да	Да	Средняя	Да	Да
Deductor Studio	Высокая	Да	Высокая	Средняя	Нет	Нет	Нет	Высокая	Нет	Да
Neural network toolbox	Низкая	Нет	Низкая	Средняя	Нет	Нет	Нет	Низкая	Нет	Да
MemBrain Neural Network	Высокая	Нет	Низкая	Низкая	Да	Да	Да	Средняя	Да	Нет
NeuroSolutions	Средняя	Нет	Средняя	Средняя	Да	Да	Нет	Средняя	Да	Нет

K_6 – использование собственных алгоритмов – принимает значения:

$$K_6 = \begin{cases} 0, \text{нет} \\ 1, \text{да} \end{cases}$$

K_7 – автоматическое формирование нейронной сети – принимает значения:

$$K_7 = \begin{cases} 0, \text{нет} \\ 1, \text{да} \end{cases}$$

K_8 – процедура импорта результатов – принимает значения:

$$K_8 = \begin{cases} 0, \text{низкая} \\ \frac{1}{2}, \text{средняя} \\ 1, \text{высокая} \end{cases}$$

K_9 – генератор исходного кода – принимает значения:

$$K_9 = \begin{cases} 0, \text{нет} \\ 1, \text{да} \end{cases}$$

K_{10} – взаимосвязь (корреляция) событий – принимает значения:

$$K_{10} = \begin{cases} 0, \text{нет} \\ 1, \text{да} \end{cases}$$

Существует наилучший вектор K^*, в котором все значения критериев соответствуют максимальным значениям. Для всех критериев это значение 1.

$$K^* = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1).$$

Для оценки качества программ вводится скалярная величина, равная Эвклидову расстоянию между наилучшим вектором и вектором критериев, полученным для *i*-й оцениваемой программы:

$$K^i = \left(K_1^i, K_2^i, K_3^i, K_4^i, K_5^i, K_6^i, K_7^i, K_8^i, K_9^i, K_{10}^i\right).$$

Эвклидово расстояние для *i*-й программы рассчитывается по формуле (1).

$$P^i = \sqrt{\sum_{j=1}^{10} \left(K_j^* - K_j^i\right)^2}. \qquad (1)$$

Программу, для которой расстояние до наилучшего вектора окажется наименьшим, можно считать наиболее рациональной программой для интеллектуального анализа данных.

Предложенная формальная модель поможет выбрать наилучшую программу интеллектуального анализа событий. Если требования к анализируемым программам изменятся, то изменив значения в наилучшем векторе K^*, можно также прийти к верному решению. Таким образом, разработанная формальная модель оценки является универсальной и эффективной.

СПИСОК ЛИТЕРАТУРЫ

1. Варлатая, С. К. Анализ угроз нарушения информационной безопасности информационных систем, существующие модели и методы противодействия компьютерным атакам / С. К. Варлатая, А. В. Кирьяненко // Актуальные проблемы технических наук в России и за рубежом : сб. науч. тр. по итогам Междунар. науч.-практ. конф., г. Новосибирск, 10 февр. 2015 г. Вып. II. – Новосибирск : Инновационный центр развития образования и науки, 2015. – С. 9–13.

2. Никулин, А. Н. Аналитическая платформа «Дедуктор» – применение в информационных системах экономики: методические указания / А. Н. Никулин, И. В. Чернышев. – Ульяновск : Изд-во УГУ, 2012. – 37 с.

3. Туровский, А. Я. Сравнительный анализ программных пакетов для работы с искусственными нейронными сетями / Я. А. Туровский, С. Д. Кургалин, А. А. Адаменко // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2016. – № 1. – С. 161–168.

REFERENCES

1. Varlataya S.K., Kiryanenko A.V. Analiz ugroz narusheniya informatsionnoy bezopasnosti informatsionnykh sistem, sushchestvuyushchie modeli i metody protivodeystviya kompyuternym atakam [Analysis of Threats to Information Security Violations, Existing Models and Methods of Countering Computer Attacks]. *Aktualnye problemy tekhnicheskikh nauk v Rossii i za rubezhom: Sbornik nauchnykh trudov po itogam mezhdunarodnoy nauchno-prakticheskoy konferentsii* [Current Problems of Technical Sciences in Russia and Abroad. Proceedings of the International Scientific and Practical Conference]. Novosibirsk, Innovation Centre for Education and Science, 2015, iss. 2, pp. 9-13.

2. Nikulin A.N., Chernyshev I.V. *Analiticheskaya platforma «Deduktor» – primenenie v informatsionnykh sistemakh ekonomiki: metodicheskie ukazaniya*

[Analytical Platform Deductor: Application in Economic Information Systems: Guidelines]. Ulyanovsk, Izd-vo UGU, 2012. 37 p.

3. Turovskiy A.Ya., Kurgalin S.D., Adamenko A.A. Sravnitelnyy analiz programmnykh paketov dlya raboty s iskusstvennymi neyronnymi setyami [Comparative Analysis of Software Packages for Work with Artificial Neural Networks]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyy analiz i informatsionnye tekhnologii*, 2016, no. 1, pp. 161-168.

THE DEVELOPMENT OF FORMAL MODEL FOR RESEARCH OF INFORMATION SYSTEM'S DATA MINING PROGRAMS

Arina Valeryevna Nikishova

Candidate of Sciences (Engineering), Associate Professor,
Department of Information Security,
Volgograd State University
arinanv@mail.ru, infsec@volsu.ru
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Svetlana Vladimirovna Mikhalchenko

Student,
Volgograd State University
infsec@volsu.ru
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Abstract. According to statistics, the number of samples of new attacks against information systems is increasing due to inability to detect new patterns and the lack of modern attack detection systems. To resolve this issue, we implement intelligent data analysis to detect attacks. There are many data mining programs, so it is important to choose the best program. The authors investigate the problem of information security from the viewpoint of new attacks, the programs for mining of information system's events. The criteria for their evaluation have been formulated. Besides, the formal model for the study of programs for mining of information system's events has been developed. The proposed formal model will help to choose the best program for event mining. If the requirements for the analyzed programs change, then changing the values in the best vector can also become a right solution. Thus, the developed formal evaluation model is universal and effective.

Key words: information security, mining, STATISTICA Automated Neural Networks, Deductor Studio, Neural network toolbox, MemBrain Neural Network, Neuro Solutions.