УДК 004.853

05.00.00 Технические науки

ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНЫХ ПАРАМЕТРОВ ФУНКЦИОНИРОВАНИЯ ИСКУССТВЕННОЙ ИММУННОЙ СИСТЕМЫ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ОБНАРУЖЕНИЯ ПОЛИМОРФНЫХ ВИРУСОВ

Частикова Вера Аркадьевна
к.т.н., доцент
РИНЦ SPIN-код: 4525-0290
chastikova_va@mail.ru
Кубанский государственный технологический университет, Краснодар, Россия

Берёзов Максим Юрьевич
магистрант
maxberezov@gmail.com
Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия

Настоящая статья посвящена исследованию параметров разработанной искусственной иммунной системы для решения задачи обнаружения полиморфных вирусов. Целью является определение такого вектора параметров иммунной системы, который бы обеспечивал минимальное количество ошибок первого рода на репрезентативной выборке данных, минимальное количество ошибок второго рода и максимальный процент обнаружения полиморфных вирусов, то есть правильной классификации их как вредоносного кода, по отношению к любому теоретически возможному вектору параметров искусственной иммунной системы. Отличительной чертой исследуемой искусственной иммунной системы является применение класса генетических алгоритмов, которые обеспечивают более эффективное обучение детекторов. Среди настраиваемых параметров работы системы выделены: алгоритм определения меры близости детектора и патогена, который может быть реализован путем определения расстояния по Левенштейну, либо методом смежных бит; а также метод реализации оператора кроссинговера, метод реализации оператора мутации, метод реализации оператора селекции, алгоритм определения меры близости строк детекторов. Кроме этого, в статье рассматривается целесообразность использования распределенной сети из нескольких узлов, на каждом из которых будет функционировать иммунная система, обменивающаяся данными с другими узлами сети. В результате исследований был получен набор оптимальных параметров, при которых система достигает максимальной точности распознавания полиморфных вирусов

UDC 004.853

Technical sciences

DETERMINATION OF THE OPTIMAL OPERATING PARAMETERS OF ARTIFICIAL IMMUNE SYSTEM TO SOLVE THE DETECTION PROBLEM OF POLYMORPHIC VIRUSES

Chastikova Vera Arkadyevna
Cand.Tech.Sci., associate professor
RSCI SPIN-code: 4525-0290
chastikova_va@mail.ru
Kuban State Technological University, Krasnodar, Russia

Berezov Maxim Yurievich
master student
maxberezov@gmail.com
Saint-Petersburg national research university of information technologies, mechanics and optics, Saint-Petersburg, Russia

This article is dedicated to the study of the parameters of the artificial immune system for solving the polymorphic viruses' detection problem. The goal is to define a vector of the immune system parameters that would ensure the minimum number of errors of the first kind, the minimum number of errors of the second kind and the maximum percentage of polymorphic viruses' detection. That is, the most accurate classification of them as a malicious code, in relation to any theoretically possible vector of parameters of the artificial immune system. A distinctive feature of the studied artificial immune system is the use of a class of genetic algorithms that provide more efficient training of detectors. The configurable parameters of the system are: the algorithm for determining the proximity of the detector and the pathogen, which can be realized by determining the Levenshtein distance or by the method of adjacent bits; as well as the method of implementing the crossing-over operator, the method of implementing the mutation operator, the method of implementing the selection operator, the algorithm for determining the proximity of the detector lines. In addition, the article considers the expediency of using a distributed network of several nodes, each of which will have an immune system that will exchange data with other nodes of the network. As a result of the research, a set of optimal parameters was obtained in which the system achieves the maximum accuracy of recognition of polymorphic viruses

Ключевые слова: ИСКУССТВЕННАЯ ИММУННАЯ СИСТЕМА, ОБНАРУЖЕНИЕ ВИРУСОВ, БИОИНФОРМАТИКА, ГЕНЕТИЧЕСКИЙ АЛГОРИТМ

Keywords: ARTIFICIAL IMMUNE SYSTEM, VIRUS DETECTION, BIOINFORMATICS, GENETIC ALGORITHM

Doi: 10.21515/1990-4665-128-031

В данной статье проводится анализ зависимости количества ошибок первого рода, количества ошибок второго рода и точности классификации полиморфных вирусов от вектора входных параметров для искусственной иммунной системы, методика и принципы работы которой были предложены в статье [1]. Для повышения эффективности работы иммунной системы используется класс генетических алгоритмов [2,5]: применение концепций генетических операторов позволяет увеличить скорость обучения детекторов. Ожидаемые результаты функционирования такой системы чувствительны к любым изменениям входного многомерного вектора, поэтому задача нахождения локального экстремума целевой функции (точность классификации полиморфных вирусов) при имплементации искусственной иммунной системы выходит на первый план [3].

В качестве исследуемых параметров функционирования системы были выбраны: методы реализации кроссинговера, селекции, мутации, метод определения близости детектора и патогена, мера близости строк детектора и мера близости строк детектора и патогена, а также целесообразность реализации многоузловой распределенной искусственной иммунной системы. Тестирование происходило на репрезентативной выборке вирусов, процесс тестирования состоял в осуществлении новых обучений искусственной иммунной системы, когда на вход подавался вектор входных параметров, отличающихся значениями только в одной позиции.

1. Определение оптимальной меры близости строк детектора

Данный параметр показывает, в каких пределах два детектора будут считаться схожими. В зависимости от свободных ресурсов система настраивается таким образом, чтобы два детектора идентифицировались различными при разных значениях параметра. Если ресурсов мало, то большое количество детекторов будут считаться подобными и отбрасываться при генерации. Таким образом, увеличивается гибкость, масштабируемость системы. На рисунке 1 показана зависимость точности распознавания вирусов от меры соответствия строк детектора.

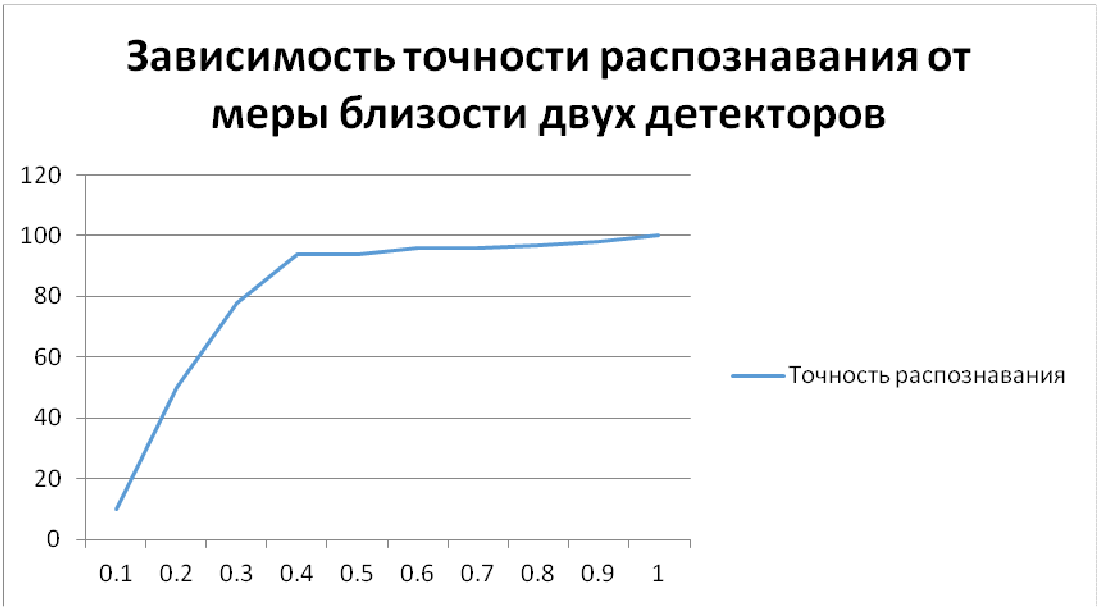

Рисунок 1 - Зависимость точности распознавания от меры соответствия строк детектора

Из полученных данных следует, что точность распознавания является кумулятивной функцией, оптимальное значение: 0.4. При указанном значении достигается 98% точность распознавания по отношению к асимптотическому максимуму, однако на заданном уровне меры близости не наблюдается резкого экспоненциального роста вычислительных ресурсов системы.

2. Определение оптимальной меры близости строк детектора и патогена

Оптимизацию данного параметра можно свети к оптимизации двух зависимых монотонных функций, первая функция точности распознавания является невозрастающей, а вторая функция отношения неложных срабатываний является неубывающей, обе функции зависят от определяемого параметра, который находится как координата X точки пересечения двух графиков. На рисунке 2 представлены функция точности распознавания и функция, показывающая количество неложных срабатываний, на основе значений которых определяется оптимальная мера близости строк детектора и патогена. По оси X - мера соответствия строк, по оси Y - процентная точность распознавания.

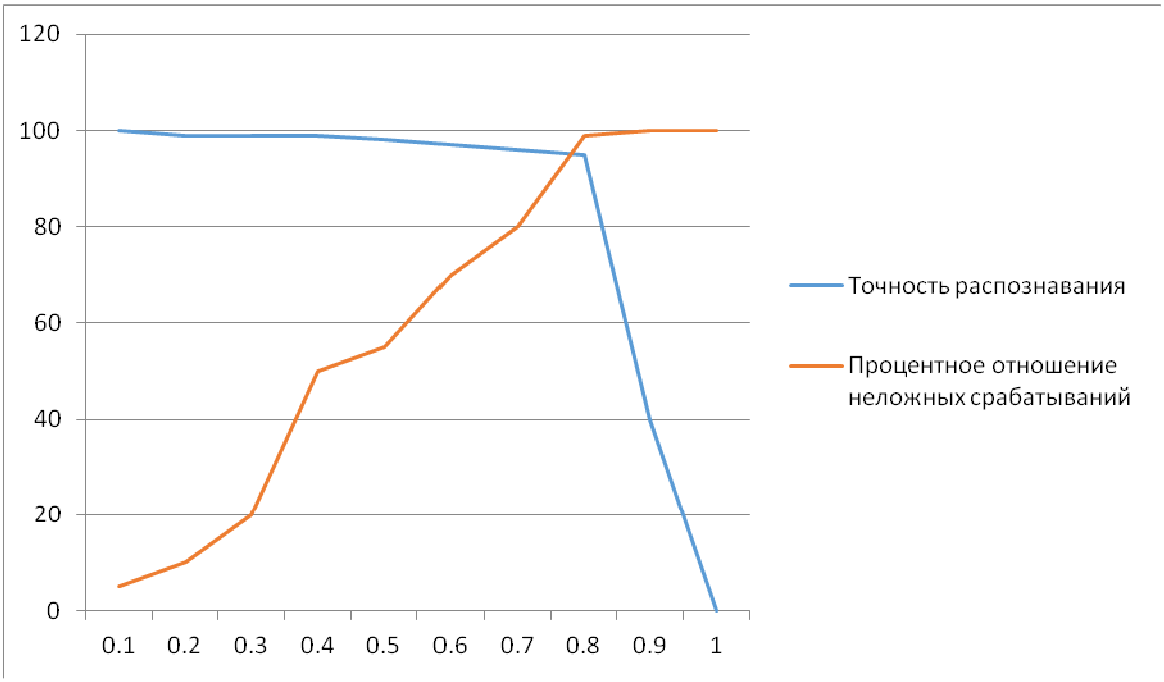

Рисунок 2 - Определение оптимального значения соответствия строки детектора патогену

Из графика видно, что при увеличении значения по оси X падает точность распознавания, но увеличивается процентное отношение неложных срабатываний. Найдя точку пересечения двух графиков, получим, что r = 0.85.

3. Выбор оптимального метода определения меры близости строк

Были реализованы два базовых алгоритма – определение расстояния Левенштейна и метод смежных бит. На рисунке 3 представлены результаты работы указанных подходов, по оси X - количество детекторов, по оси Y - процентная точность распознавания.

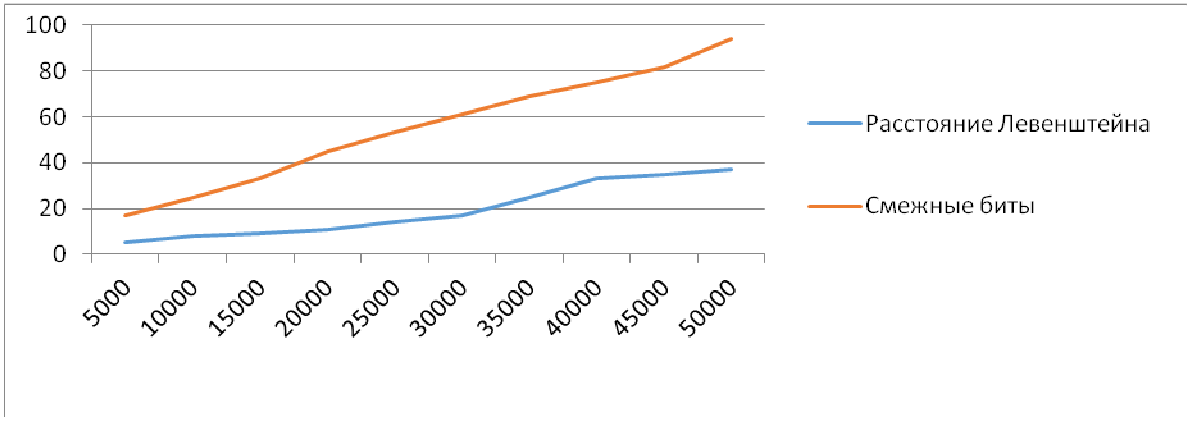

Рисунок 3 – Зависимость точности распознавания вирусов от метода определения меры близости строк

Очевидно, что подсчет числа смежных бит показывает большую точность классификации, что делает его оптимальным методом для задачи сравнения двух строк в поставленной задаче.

4. Определение оптимального метода реализации операторов кроссинговера, мутации и селекции

В разработанной искусственной иммунной системе [1] возможна реализация оператора кроссинговера с помощью методов [2,5]:

- случайной маски;
- одноточечный кроссинговер.

Оператор мутации представлен следующими методами:

- одноточечная мутация;
- многоточечная мутация;
- инверсия бит;
- инверсия относительно точки разреза.

Для реализации селекции используются подходы:

- турнирная селекция;

- элитарная селекция;

- метод колеса рулетки.

Результаты экспериментов и сравнительного анализа для всех трех методов представлены в таблицах 1-3. Замеры показаний происходили с дискретным шагом в 5000 детекторов.

Таблица 1 – Сравнение методов реализации кроссинговера

Количество детекторов	Процентная точность распознавания на основе метода случайной маски	Процентная точность распознавания на основе метода одноточечного кроссинговера
5000	20	14
10000	27	21
15000	37	30
20000	44	37
25000	55	49
30000	65	58
35000	72	69
40000	79	75
45000	86	87
50000	94	94

Для большого количества детекторов оба метода дают одинаковый результат и по точности распознавания, и по проценту ложных срабатываний, однако при меньшем количестве детекторов метод случайной маски дает лучшие результаты.

Таблица 2 – Сравнение методов реализации мутации

	Процентная точность распознавания			
Количество детекторов	Одноточечная мутация	Многоточечная мутация	Инверсия бит	Инверсия относительно точки
5000	20	21	22	21
10000	27	29	28	27
15000	37	39	41	42
20000	44	46	45	48
25000	55	57	59	58
30000	65	66	65	69
35000	72	74	78	77
40000	79	81	79	83
45000	86	87	87	89
50000	92	94	94	95

Все четыре метода мутации показывают примерно одинаковые результаты в обоих тестах. Наиболее простым и гибким считается метод инверсии бит относительно точки разреза, который обладает большей точностью при увеличении популяции детекторов.

Таблица 3 – Сравнение методов реализации селекции

	Процентная точность распознавания		
Количество детекторов	Метод колеса рулетки	Элитарная селекция	Турнирная селекция
5000	20	17	13
10000	27	24	22
15000	37	33	30
20000	44	42	40
25000	55	53	51
30000	65	60	58
35000	72	67	65
40000	79	73	70
45000	86	81	79
50000	94	90	88

Наилучшие результаты показывает метод колеса рулетки: процентная точность данного подхода при любом количестве детекторов выше, чем у смежных методов.

Таким образом, оптимальными реализациями представленных генетических операторов являются метод случайно маски для кроссинговера, метод инверсии бит относительно точки разреза для мутации и метод колеса рулетки для оператора селекции.

5. Определение целесообразности использования распределенной сети из нескольких узлов

Перспективным подходом увеличения точности классификации в распределенной сети из нескольких узлов является использование искусственной иммунной системы на каждом узле и обмен базами данных детекторов через набор протоколов TCP/IP.Точность классификации при использовании распределенной сети из нескольких узлов по сравнению с одним узлом при проведении экспериментов увеличилась на один процент; для распределенной сети на графике количество детекторов по оси X равно

среднему количеству детектору на одном узле. На рисунке 4 приведена точность обнаружения для одного узла и распределенной сети из пяти узлов.

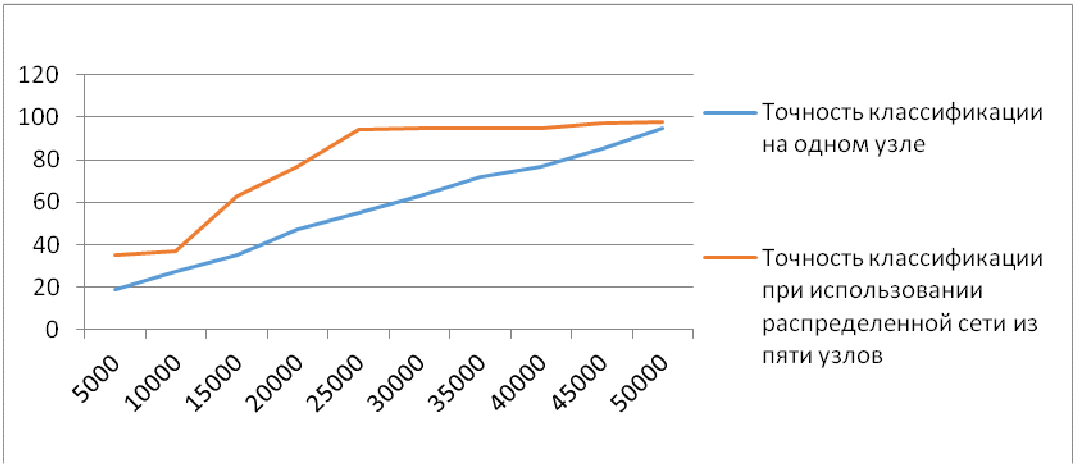

Рисунок 4 – Точность обнаружения для одного узла и распределенной сети из пяти узлов

Таким образом, использование распределенной сети из нескольких узлов является целесообразным, точность классификации увеличивается при одинаковой нагрузке на узел.

6. Оптимальные параметры работы искусственной иммунной системы

На основе вышеприведенных данных сравнительного анализа оптимальными параметрами функционирования разработанной искусственной иммунной системы являются:

– мера близости строк детектора и патогена: 0,85;

– мера близости строк детекторов: 0,4;

– метод определения меры близости строк: метод смежных бит;

– метод реализации мутации строки детектора: метод мутации относительно точки разреза;

– метод реализации кроссинговера строки детектора: метод случайной маски;

– метод реализации селекции: метод турнирной селекции;

– целесообразность использования распределенной сети иммунных систем из нескольких узлов: да.

Для такой системы точность распознавания вирусов представлена на рисунке 5, где по оси X указано количество используемых детекторов, а по оси Y - процентная точность классификации.

Рисунок 5 - Точность распознавания для ИИС с оптимальными параметрами работы

При большом количестве детекторов удалось добиться точности распознавания приблизительно в 96 %, дальнейший рост точности в зависимости от количества детекторов носил очень медленный характер, что говорит о наличии горизонтального асимптотичного предела для данной функции. Стоит отметить, что система, имплементируемая в соответствии с полученным входным многомерным вектором параметров, будет являться высокоточным классификатором зловредного полиморфного кода.

Литература

1. Частикова В.А., Берёзов М.Ю. Методика обнаружения полиморфных вирусов на основе искусственных иммунных систем и генетических алгоритмов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2016. № 124. С. 744-755.

2. Частикова В.А. Идентификация механизмов реализации операторов генетического алгоритма в экспертных системах продукционного типа //Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2012. № 75. С. 308-320.

3. Частикова В.А., Картамышев Д.А. Искусственные иммунные системы:

основные подходы и особенности их реализации //Научные труды Кубанского государственного технологического университета. 2016. № 8. С. 193-208.

4. Частиков А.П., Белов Д.Л. Структура регенеративной экспертной системы // Материалы VII Международной научно-практической конференции «Инновационные процессы в высшей школе». 2001. С. 107-108.

5. Симанков В.С., Частикова В.А. Генетический поиск решений в экспертных системах. - Краснодар, 2008.

6. Частиков А.П., Малыхина М.П., Урвачев П.М. Анализ распознавания паттернов нейросетевыми методами // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 98. С. 457-467.

7. Малыхина М.П., Бегман Ю.В. Оценка эффективности гибридизации интеллектуальных методов на примере нейросетевой экспертной системы на основе прецедентов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 86. С. 253-262.

8. de Castro Leandro N. Artificial Immune Systems: A New Computational Intelligence Approach. - Springer, 2002.

9. Kephart, J. O. (1994). "A biologically inspired immune system for computers". Proceedings of Artificial Life IV: The Fourth International Workshop on the Synthesis and Simulation of Living Systems: 130-139, MIT Press.

10. D. Dasgupta (Editor), Artificial Immune Systems and Their Applications, Springer-Verlag, Inc. Berlin, January 1999.

References

1. Chastikova V.A., Berjozov M.Ju. Metodika obnaruzhenija polimorfnyh virusov na osnove iskusstvennyh immunnyh sistem i geneticheskih algoritmov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2016. № 124. S. 744-755.

2. Chastikova V.A. Identifikacija mehanizmov realizacii operatorov geneticheskogo algoritma v jekspertnyh sistemah produkcionnogo tipa //Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2012. № 75. S. 308-320.

3. Chastikova V.A., Kartamyshev D.A. Iskusstvennye immunnye sistemy: osnovnye podhody i osobennosti ih realizacii //Nauchnye trudy Kubanskogo gosudarstvennogo tehnologicheskogo universiteta. 2016. № 8. S. 193-208.

4. Chastikov A.P., Belov D.L. Struktura regenerativnoj jekspertnoj sistemy // Materialy VII Mezhdunarodnoj nauchno-prakticheskoj konferencii «Innovacionnye processy v vysshej shkole». 2001. S. 107-108.

5. Simankov V.S., Chastikova V.A. Geneticheskij poisk reshenij v jekspertnyh sistemah. - Krasnodar, 2008.

6. Chastikov A.P., Malyhina M.P., Urvachev P.M. Analiz raspoznavanija patternov nejrosetevymi metodami // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 98. S. 457-467.

7. Malyhina M.P., Begman Ju.V. Ocenka jeffektivnosti gibridizacii intellektual'nyh metodov na primere nejrosetevoj jekspertnoj sistemy na osnove precedentov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 86. S. 253-262.

8. de Castro Leandro N. Artificial Immune Systems: A New Computational Intelligence Approach. - Springer, 2002.

9. Kephart, J. O. (1994). "A biologically inspired immune system for computers". Proceedings of Artificial Life IV: The Fourth International Workshop on the Synthesis and Simulation of Living Systems: 130-139, MIT Press.

10. D. Dasgupta (Editor), Artificial Immune Systems and Their Applications, Springer-Verlag, Inc. Berlin, January 1999.