Era of Advanced Persistent Threats", IEEE Security & Privacy, Vol.12, №5, IEEE, 2014

3. И. В. Котенко, М. В. Степашкин. Обманные системы для защиты информационных ресурсов в компьютерных сетях // Труды СПИИРАН, Вып. 2, т. 1. — СПб.: СПИИРАН, 2004.

4. Bringer M.L., Chelmecki, C.A., Fujinoki H A Survey: Recent Advances and Future Trends in Honeypot Research. // International Journal of Computer Network & Information Security. 2012. №4.

5. В.Г.Олифер, Н.А.Олифер. Компьютерные сети. Принципы, технологии, протоколы - СПб: Питер, 2001.

6. Морозов Л.М., Петухов Г.Б., Сидоров В.Н. Методологические основы теории эффективности: Учебное пособие. – Л.: ВИКИ им. А.Ф, Можайского, 1982. – 236с.

М.Е. Бурлаков

О НЕКОТОРЫХ МОДЕЛЯХ ОПТИМИЗАЦИИ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ГЕНЕТИЧЕСКИМИ АЛГОРИТМАМИ

(Самарский государственный университет)

В настоящее время в информационной среде все больше находят свое применение искусственные нейронные сети (ИНС). Выделяют множество таковых областей: системы обнаружения вторжений, системы предотвращения вторжений, интеллектуальные экспертные системы, системы прогнозирования [1] и т.д. Для каждой реализации той или иной системы, проектируется и применяется конкретное решение, оптимально отвечающее поставленной задаче. Однако вопрос повышения эффективности уже выбранного решения на основе искусственной нейронной сети является крайне актуальным. Не существует общего алгоритма подбора оптимальных параметров искусственной нейронной сети (вес нейрона, общая топология ИНС, функция активации). Одним из методов, предложенных для решения данной задачи, является применение генетических алгоритмов (ГА).

Генетические алгоритмы

Генетические алгоритмы – одно из направлений исследований в области искусственного интеллекта, занимающееся созданием упрощенных моделей эволюции живых организмов для решения задач оптимизации [1]. Генетические алгоритмы – эвристический алгоритм поиска, используемый для решения задач оптимизации и моделирования путём случайного подбора, комбинирования и вариации искомых параметров с использованием механизмов, аналогичных естественному отбору в природе [2]. Другими словами ГА - представляет собой адаптивный поисковый метод, который основан на селекции лучших элементов в популяции, подобно эволюционной теории Ч. Дарвина.

Эвристический алгоритм поиска (эволюционный поиск) - с точки зрения преобразования информации это последовательное преобразование одного конечного нечеткого множества промежуточных решений в другое. Само преобразование можно назвать алгоритмом поиска, или генетическим алгоритмом [3].

ГА осуществляют поиск баланса между эффективностью и качеством решений за счет «выживания сильнейших альтернативных решений», в неопределенных и нечетких условиях. Выделяют большое множество моделей ГА, целью которых является оптимизация и/или решение поставленной задачи.

У генетических алгоритмов, как у всякой замкнутой системы существует базис операций. Выделяют три основных вида операции:

1. Селекция – процесс, при котором объект информационной структуры с лучшими характеристиками в рамках конкретного набора аналогичных объектов выделяется из общего множества аналогичных объектов и используется для создания нового множества. Например, в антивирусной среде объектом послужил бы набор байтов (сигнатура), тогда как селекция описывала бы процесс, при котором данная сигнатура наилучшим образом детектировала угрозу.
2. Скрещивание. В данном процессе выделяет пара объектов, в результате скрещивания которых образуются новые объекты следующего поколения с отличными от родительских характеристиками. Скрещивание на примере той же сигнатуры означает, что из двух объектов, успешно и максимально детектирующих угрозу предполагается получение третьей сигнатуры, способной решать аналогичную задачу.
3. Мутация. Операция, при которой объект подвергается внутренней корректировки со стороны системы. Выделяют интенсивность мутации – скорость, с которой система может воздействовать на объект информационного множества.

Иллюстрация основных операций генетических алгоритмов представлена на рисунке 1.

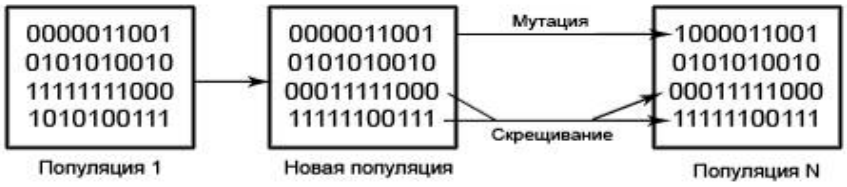

Рис. 1. Основные операции генетического алгоритма

Применение генетических алгоритмов основываются на принципах естественной селекции: в ходе развития набора элементов или объектов (популяции) из поколения в поколение «выживает» наиболее эффективный. Базис объектов ГА в информационных системах носит бинарный характер.

Модели работы генетических алгоритмов

В зависимости от комбинации и работы операций селекции, скрещивания и мутации, выделяют несколько моделей работы ГА:

***Parallel genetic algorithm* модель**

В условиях динамичного развития многопроцессорных систем – на практике это наиболее простой в реализации набор алгоритмов. Выделяют 3 типа *PGA*:

1. *Global single-population master-slave (GSPMS)*. Данный тип характеризуется применимостью в рамках одного множества бинарных объектов. Уровень *Master* (главный уровень) отвечает за хранение всего множества, выполнение базисных операций и распределение полученных объектов среди новых созданных поколений. *Slave* (подчиненный уровень) выполняет операцию вычисления эффективности полученных информационных объектов.
2. *Single-population fine-grained (SPFG).* Тип *PGA* модели работающий только в рамках одного множества. Замкнутая система, в котором результат операций есть объект, принадлежащий данному бинарному множеству (популяции). Наиболее оптимальным выглядит использование *SPFG* в многопроцессорных системах.
3. *Multiple-population coarse-grained (MPCG)* — характеризуется наличием в одном бинарном множестве, которые обмениваются между собой объектами в ходе работы алгоритма. Операции мутации и скрещивания на вход могут принимать более двух объектов. *MPCG* на данный момент активно исследуется, так как есть ряд вопросов касающихся неоднозначности поведения системы, например, при объединении или миграции нескольких множеств [4].

Плюсы:

- Скрещивание, селекция и мутация – параллельные процессы;
- Детерминированная мощность множества объектов, за счет добавления наиболее эффективных элементов и исключения наиболее неэффективных. Мощность множества не регламентирована;
- Возможность применения в многопроцессорных системах, что значительно ускоряет результат работы;
- Основная область применения задача поиска аномалий (например, поиск ложных подписей ЭЦП, сенсорная защита, системы анализа трафика).

Минусы:

- Не определен эффективный алгоритм мутации двух объектов из разных классов распознавания (например, класс распознавания одного вируса и другого).

Cross generational elitist selection, Heterogeneous recombination and Cataclysmic mutation.

Модель, при которой получение нового объекта связано со скрещиванием наиболее приспособленных к распознаванию сигнатур объектов. *CHC* алгоритм произвольно скрещивает объекты одного множества и одного поколения. Воспроизводство нового поколения возможно только у тех объектов, которые от-

личаются друг от друга на некоторое количество бит (т.н. порог воспроизводства). С тем условием, что данные объекты отличаются друг от друга не незначительное количество бит. После получения новых объектов, создается новое множество, куда они включаются. Одинаковые объекты удаляются при перемещении. Модель наиболее часто используется для небольших множеств.

В случае отсутствия потомков у скрещиваемых объектов порог воспроизводства снижают на некоторое конечное значение. Оператор скрещивания выполняет перекрестное объединение строк двух объектов с последующим случайным исключением половины бит, которые отличаются от первоначальных индивидов.

Важное условие – отсутствие операции мутации на этапе рекомбинации. В случае если при скрещивании объектов потомства нет, а порог воспроизводства упал до 0, *CHC* алгоритм предусматривает введение новых объектов в популяцию путем перезапуска механизма воспроизведения, называемого предельной мутацией.

Предельная мутация использует наиболее эффективные индивиды из старого множества в качестве шаблона для переинициализации следующего множества. Новое множество включает в себя одну копию строки шаблона, тогда как остальные строки генерируются путем мутации некоторого процента битов второй строки (от второй особи). Уровень мутации битов может доходить до 35% [5].

Плюсы:

- Наличие механизма самокорректировки.
- Основная область применения задача классификации (например, детектирование Спама, решение задачи приоритезации потоков в системах электронного документооборота).

Минусы:

- Необходимость подбора порога воспроизводства и уровня мутации.
- сложность вычислений при большом количестве объектов в множестве.

Метод *Hierarchical Genetic Algorithm and Neural Network*

HGAANN используется для оптимизации искусственных нейронных сетей. *HGANN* имеет иерархическую структуру, в которой каждая хромосома состоит из многоуровневых генов. Под генами, в данном случае, понимается набор нейронов ИНС (рисунок 2). В модели, как правило, выделяют 3 вида генов [6]:

1. *LCG* (*Layer Control Gen*) – ген по управлению слоями, отвечает за обработку поступающей информации и последующую передачу на уровень *NCG*;
2. *NCG* (*Neuron Control Gen*) – ген, содержащий в себе многослойный модуль обработки данных ИНС;
3. *CG* (*Connection Gen*) – ген, отвечающий за отображение полученной информации и последующего взаимодействия с внешними системами (значение ИНС).

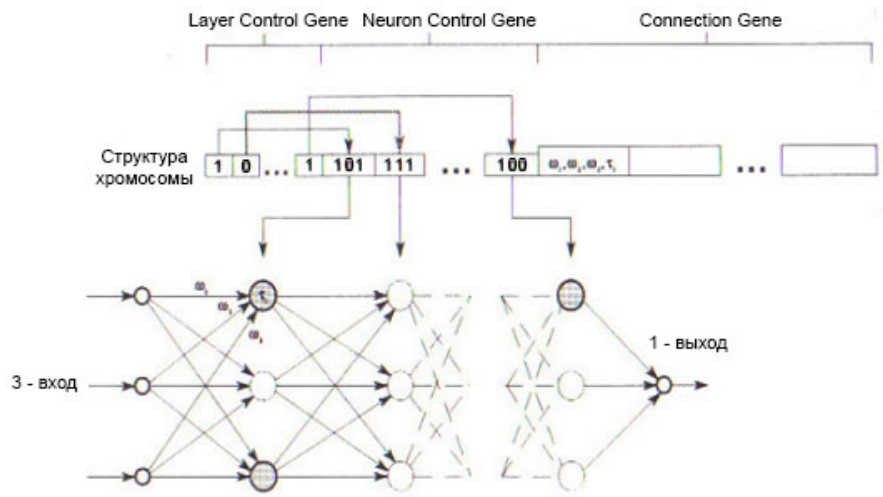

Рис. 2. Схема *HGANN*

Каждый ген состоит из двух частей: контрольной и коммутационной. Контрольная часть - есть нейрон и активационная функция ω, а коммуникационный ген – вес нейрона и его бинарный базис. Неактивные гены, как правило, уничтожаются, при создании нового поколения объектов. Основная задача данной конструкции: поддержка баланса между имеющейся базой знаний ИНС и дальнейшей ее актуализацией. Разовое изменение какого-либо параметра на ранних слоях влечет существенные изменения в полученных результатах работы *HGANN*.

Структурная схема работы алгоритма *HGA* представлена на рисунке 3.

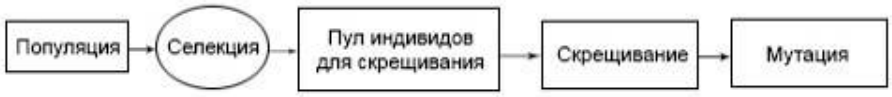

Рис. 3. Блок-диаграмма *HGA*

В пул объектов для скрещивания попадают все представители множества объектов, контрольная часть генов (подмножества множества бит данного объекта) которых отвечают условиям оптимальной выборки.

Опишем основные шаги работы метода *HGANN*:

1. Получение новой популяции после работы *HGA* алгоритма стадии мутации;
2. Сборка необходимых нейронов, отвечающих заданным параметрам (вес, функция активации, топология) согласно индивидам из новой популяции
3. ИНС, обучение и тренировка;
4. Оценка полученных результатов работы ИНС, в случае получения неоптимальных параметров генерация новой популяции;
5. Замена старой популяции новой и передача на вход *HGA* алгоритму.

Общая блок-диаграмма метода *HGANN* представлена на рисунке 4.

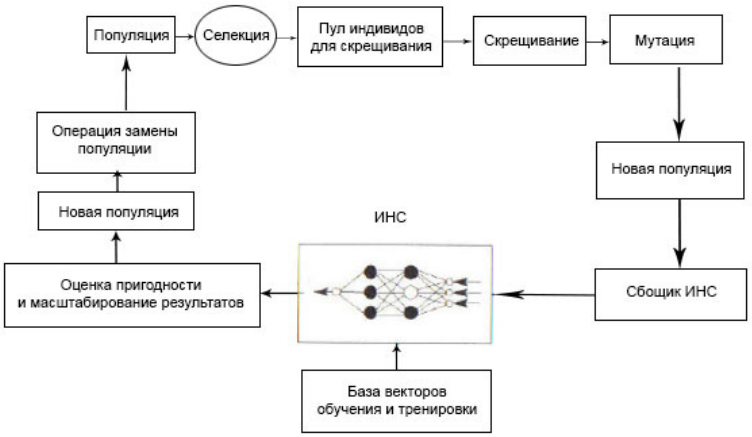

Рис. 4. Блок-диаграмма *HGANN*

Заключение

Продемонстрированные методы и модели в ходе практического анализа показали возможность оптимизации ИНС. Кроме того, выбор ГА характеризовался решаемой ИНС информационной задачи. Однако вопрос повышения эффективности ИНС нелинеен и связан, прежде всего, с попытками увеличения эффективности работы каждой из ее компонент: топологии, весовых параметров, функций активации. Рассмотренные *HGANN*, *PGA* и *CHC* модели показывают общую тенденцию применимости генетических алгоритмов в вопросах оптимизации искусственных нейронных сетей в информационных системах.

Литература

1. Девятых Д. В., Гергет О. М., Михаленко И. В. Применение искусственных нейронных сетей для прогнозирования развития перинатального поражения нервной системы. – Известия Волгоградского государственного технического университета : межвуз. сб. науч. ст. № 8(111) / ВолгГТУ. – Волгоград : ИУНЛ ВолгГТУ, 2013. – С. 77.

2. Генетические алгоритмы [Электронный ресурс]. – 2012. – Режим доступа: http://www.aiportal.ru/articles/genetic-algorithms/genetic-algorithms.html

3. Генетические алгоритмы [Электронный ресурс]. – 2013. – Режим доступа: http://ru.wikipedia.org/wiki/%D0%93%D0%B5%D0%BD%D0%B5%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D0%B0%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC

4. Определения и основные понятия генетических алгоритмов: Интернет-лекция в рамках научной школы Третьей конференции молодых ученых [Электронный ресурс]. – 2013. – Режим доступа: http://faculty.ifmo.ru/csd/files/kureichik_v_m_lebedev_b_k_internet_lecture.pdf

5. Erick Cantú-Paz A Survey of Parallel Genetic Algorithms // Department of Computer Science and Illinois Genetic Algorithms Laboratory University of Illinois at Urbana-Champaign, 2013. – С. 5.

6. Darrell Whitley Genetic Search for Feature Subset Selection: A Comparison Between CHC and GENESIS // Department of Computer Science Colorado State University Fort Collins, Colorado 80523 USA, 2012. – С. 7.

7. Dr. M.V. Siva Prasad An Intrusion Detection System Architecture Based on Neural Network and Genetic Algorithms // Principal Anurag Engineering College, 2013. – С. 6.

Л.Ф. Зиангирова, Т.И. Саттаров

СТЕГАНОАНАЛИТИЧЕСКИЕ МЕТОДЫ ДЛЯ ПЕРЕСЫЛАЕМЫХ КОНТЕЙНЕРОВ

(Башкирский государственный педагогический университет им. М. Акмуллы,
Уфимский государственный нефтяной технический университет)

Появление методов скрытой передачи данных посредством пересылаемых контейнеров привело к развитию стеганографического программного обеспечения. Пересылаемый контейнер – это объект в виде текста, архива, фотографии, звукового файла, видео файла и других объектов, пересылаемых через электронную почту, социальные сети, форумы, системы мгновенного обмена сообщениями, а также копирование на внешние носители.

Существуют различные решения для защиты сети предприятия от утечки из нее конфиденциальной информации. Данный класс решений получил название DLP (Data Leakage Prevention). DLP-система предназначена для контроля информационных потоков, защиты конфиденциальной информации от утечки и несанкционированного распространения. Известные представители DLP-систем: Websense DSS, InfoWatch, Symantec DLP, SearchInform и др. Данные системы перехватывают весь трафик, выходящий за пределы сети предприятия, и сканируют его на наличие в нем конфиденциальных данных. Также DLP-системы сканируют всю информацию, записываемую пользователями сети на съемные носители при помощи их рабочих станций. Они способны отследить конфиденциальную информацию, передаваемую в открытом или заархивированном виде, пресечь передачу зашифрованных данных.

Стеганографические программы предотвращают способность инсайдера передать конфиденциальные данные за пределы сети предприятия способом включения битов информации в мультимедийные контейнеры, которые не запрещены для передачи.

В зависимости от используемых исходных данных методы стеганоанализа можно разделить на следующие группы:

– Методы, предназначенные для работы с конкретными заранее известными стеганографическими алгоритмами.