На правах рукописи

Ле Тхи Чанг Линь

ОПТИМАЛЬНЫЕ МНОГОЭКСПЕРТНЫЕ БИНАРНЫЕ СИСТЕМЫ ГОЛОСОВАНИЯ В МОДЕЛЬНЫХ ЗАДАЧАХ ОБНАРУЖЕНИЯ АТАК

05.13.01 – Системный анализ, управление и обработка информации
(информационные и технические системы)

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Москва – 2018

Работа прошла апробацию на кафедре «Интеллектуальные информационные системы и технологии» Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (государственный университет)»

Научный руководитель: доктор технических наук,
старший научный сотрудник,
Аведьян Эдуард Дзеронович

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт проблем управления им. В. А. Трапезникова Российской академии наук.

Защита состоится 07.12.2018 в 10:00 на заседании диссертационного совета ФРКТ 05.13.01.004 по адресу 141701, Московская область, г. Долгопрудный, Институтский переулок, д. 9.

С диссертацией можно ознакомиться в библиотеке и на сайте Московского физико-технического института (государственного университета) https://mipt.ru/education/post-graduate/soiskateli-tekhnicheskie-nauki.php

Работа представлена «17» сентября 2018 г. в Аттестационную комиссию федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (государственный университет)» для рассмотрения советом по защите диссертаций на соискание ученой степени кандидата наук, доктора наук в соответствии с п. 3.1 ст. 4 Федерального закона «О науке и государственной научно-технической политике»

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы диссертации

В последнее годы интерес к результатам исследований в области искусственного интеллекта возрастал по экспоненте. В частности, большое внимание уделяется сейчас области систем, основанных на знаниях. Из систем, основанных на знаниях, экспертные системы сейчас являются наиболее успешными. Важной особенностью экспертных систем является то, что они могут объяснить пользователю линию рассуждений, которая привела к решению проблемы или желаемому совету. Эффективность экспертных систем существенно возрастает, когда решение в такой системе принимается коллективом экспертов. Подобные системы являются многоэкспертными. Многоэкспертный подход позволяет также включать в состав экспертов не только людей, но и специализированные алгоритмы. Для объединения решений отдельных экспертов разработано несколько подходов. Один из самых распространенных и популярных походов создания многоэкспертных систем - алгоритм принятия решения большинством голосов, который имеет довольно длинную историю. Несмотря на все достоинства этого метода, таких как простота реализации и естественность его обоснования, этот метод является оптимальным только в частных случаях.

Основной задачей, которая решается в настоящей работе, было проведение глубокого анализа алгоритма принятия решения большинством голосов и построение оптимальной процедуры голосования, которая в многоэкспертных системах принятия решений может быть использована в самых различных предметных областях. Это уже делает актуальным тему диссертационной работы. В качестве предметной области, на которой иллюстрируется применение оптимального многоэкспертного подхода, выбрана область информационной безопасности.

В настоящее время в связи с бурным развитием интернета, интернета вещей (Internet of things, IoT) и больших данных (big data) проблема обеспечения безопасности информационных ресурсов становится более актуальной, чем когда-либо. Одним из важнейших инструментов обеспечения информационной безопасности служат системы обнаружения вторжений (СОВ) (Intrusion Detection System – IDS). Системы обнаружения вторжений стали важной частью сегодняшней инфраструктуры сетевой безопасности, которая может отслеживать сетевой трафик и обнаруживать возможные атаки. Существующие СОВ страдают из-за низкой точности обнаружения вторжений. Системы обнаружения вторжений, в которых реализован многоэкспертный подход, обладают как правило более качественными характеристиками, чем простые экспертные системы. Поэтому второй задачей, рассмотренной в диссертационной работе, является демонстрация на модельных примерах того, как оптимальный многоэкспертный подход может повысить качество безопасности информационных систем. Решение этой задачи также определяет актуальность диссертационного исследования.

Степень разработанности темы. Различные подходы к объединению решений отдельных экспертов в многоэкспертных системах принятия решений представлены в работах российских и иностранных ученых, таких как Л. А. Растригин, Р. Х. Эренштейн,

В. И. Городецкий, А.И. Орлов, С. В. Серебряков, В. Н. Ручкин, В. А. Фулин, Б. И. Ефимов, Р. Т. Файзуллин, А. А. Браницкий, Котенко И. В., Lam L., Suen C.Y., Kuncheva L.I., Xu L., Amari Shunichi, John von Neumann, Shannon C. E., Aburomman A. A., Reaz M. B. I. В этих работах рассматриваются алгоритмы принятия решения большинством голосов (majority voting), алгоритмы принятия решения взвешенным большинством голосов (weighted majority vote), алгоритмы принятия решения, основанные на правиле Байеса (Bayes'rule), правиле Демпстера – Шафера и другие. В работе Xu Lei и Amari Shun-ichi [2] авторы выделяют две основные задачи, которые следует решать при создании многоэкспертной системой: 1. Какой тип экспертов и какое их количество необходимы для решения конкретной задачи? 2. Как объединить результаты решений отдельных экспертов, чтобы решение многоэкспертной системы было лучше, чем решения отдельных экспертов? Несмотря на очень большое число публикаций по данной тематике, до сих пор полного ответа на оба этих вопроса пор нет.

Целью исследования является изучение основных свойств одного из самых популярных методов, применяемых в многоэкспертных системах, а именно метода принятия решения большинством голосов (majority voting), придание ему оптимальных свойств как при равной, так и при различной вероятности правильного решения каждого статистически взаимно независимого и зависимого эксперта и иллюстрация применения оптимальных многоэкспертных бинарных систем голосования (МЭБС) в модельных задачах обнаружения атак на информационные ресурсы.

Для достижения этой цели в диссертационной работе сформулированы следующие задачи:

1. Провести теоретический и компьютерный анализ алгоритма принятия решения большинством голосов.

2. На основе проведенного анализа предложить и исследовать оптимальную многоэкспертную бинарную систему голосования при равных и неравных вероятностях принятия решений отдельных экспертов.

3. При неизвестных вероятностях принятия решений отдельных статистически независимых и зависимых экспертов разработать алгоритмы метода статистических испытаний для реализации оптимальной МЭБС.

4. Показать, как оптимальные МЭБС могут быть применены для создания оптимальных систем обнаружения атак.

Объектом исследования диссертации являются МЭБС в системах принятия решений, современные системы обнаружения атак и процедуры применения оптимальных МЭБС голосования для оптимизации систем обнаружения атак.

Предметом исследования диссертации являются модели, методы, алгоритмы и программы создания оптимальных многоэкспертных бинарных систем голосования, в которых в роли экспертов выступают искусственные нейронные сети.

Методы исследования. Для решения поставленных задач используется методы теории вероятностей, математической статистики, вычислительной математики, компьютерных

методов обработки информации и моделирования, искусственных нейронных сетей, а также методы разработки приложений на языках C++, Matlab и Python.

Научная новизна результатов, полученных в диссертационной работе, состоит в следующем:

1. Установлены новые свойства алгоритма принятия решения большинством голосов при **четном** числе экспертов, характеризующие вероятности принятия правильного решения МЭБС:

   1.1. дополнен результат, приведенный в работе [1], который определяет вероятность правильного решения МЭБС $p_{\exp sys}^{(0/0)}(2m)$, когда за гипотезу H_1 голосует более половины экспертов при равной вероятности правильного решения каждого эксперта p. В [1] отмечается, что данные функции являются немонотонными при $1/3 < p < 1/2$, а их свойства зависят от значений p и m. В диссертации показано что при $1/3 < p < 1/2$ функции $p_{\exp sys}^{(0/0)}(2m)$ одноэкстремальны: сначала с ростом числа экспертов $2m$ они возрастают, достигая максимума, не превышающего 0.5, после чего начинается их убывание до нуля. В диссертации определены точки их максимума.

   1.2. проведен анализ вероятностей правильного решения МЭБС $p_{\exp sys}^{(0/0)}(2m)$, когда за альтернативную гипотезу H_0 голосует ровно половина экспертов или более половины экспертов при равной вероятности правильного решения каждого эксперта p. Показано, что в этом случае с ростом числа экспертов при p, удовлетворяющих условию $2/3 < p < 1$, функции $p_{\exp sys}^{(0/0)}(2m)$ являются монотонно возрастающими и стремящимися к 1, при $0 < p < 1/2$ - эти функции - монотонно убывающие, стремящиеся к нулю. При p, удовлетворяющих условию $1/2 < p < 2/3$, функции $p_{\exp sys}^{(0/0)}(2m)$ сначала с увеличением числа экспертов убывают, достигают минимума, после чего начинается их рост и стремление к 1, определены точки минимума этих функций.

2. Сформулирован принцип построения оптимальных МЭБС, при котором определяется число экспертов, голосующих за каждую из двух гипотез так, чтобы функционал, представляющий линейную комбинацию вероятностей правильного решения МЭБС относительно каждой из гипотез, достигал максимального значения.

3. Разработан метод статистических испытаний, позволяющий найти оптимальное решение в МЭБС как при неизвестных вероятностях правильных решений отдельных экспертов, так и при их статистической зависимости.

4. Разработана технология применения оптимальной МЭБС для создания систем обнаружения атак.

5. Исследованы возможности применения нейронной сети CMAC для обнаружения атак и применения ее в оптимальных МЭБС.

6. Разработаны компьютерные программы для моделирования МЭБС, которые использованы для исследования свойств МЭБС и применения их в модельных задачах обнаружения атак.

Практическая значимость. Результаты настоящего исследования могут быть применены в областях, в которых используются многоэкспертные системы или в задаче объединения решений отдельных экспертов, в том числе в системах обнаружения атак. Результаты диссертационного исследования внедрены в Федеральном государственном автономном научном учреждении "Центр информационных технологий и систем органов исполнительной власти" (ФГАНУ ЦИТиС) и будут в дальнейшем использоваться для обеспечения информационной безопасности сетей связи специального назначения, что подтверждено актом о внедрении.

Кроме того, результаты и программы исследования используются в учебном процессе на кафедре «Интеллектуальные информационные системы и технологии» в Федеральном государственном автономном образовательном учреждении высшего образования «Московский физико-технический институт (государственный университет)» при проведении лекций и семинаров. Прототип программы моделирования системы обнаружения атак используется в курсах «Нейроматематика» и «Теория нейронных сетей» для студентов и магистров направления «Прикладные математика и физика».

Обоснованность и достоверность результатов и выводов определяется следующим факторами: строгим доказательством полученных результатов, публикациями результатов исследования в рекомендованных Высшей аттестационной комиссией научных изданиях, практическим использованием результатов диссертационной работы, подтвержденным актом о внедрении.

Личный вклад соискателя состоит в непосредственном участии в разработке моделей, математических методов, алгоритмов и программного обеспечения, проведении исследовательских испытаний, апробации результатов исследования, обработке баз данных, подготовке основных публикации по выполненной работе.

Апробация результатов работы. Материалы диссертационной работы были доложены и обсуждены на следующих международных конференциях: IX Международная научно - практическая конференция «Логистика и экономика ресурсоэнергосбережения в промышленности» 9-11 ноября 2015, Смоленск; Пятнадцатая национальная конференция по искусственному интеллекту КИИ-2016, 3-7 октября 2016, Смоленск;

XIV Всероссийская научная конференция «Нейрокомпьютеры и их применение», 15 марта 2016, Москва; III Международная конференция «Инжиниринг и телекоммуникации» – En&T, 29-30 ноября 2016, Долгопрудный; XV Всероссийская научная конференция «Нейрокомпьютеры и их применение», 14 марта 2017, Москва; 60-я научная конференция МФТИ, 20-26 ноября 2017, Долгопрудный; IV Международная конференция «Инжиниринг и телекоммуникации» – En&T, 29-30 ноября 2017, Долгопрудный; XVI Всероссийская научная конференция «Нейрокомпьютеры и их применение», 13 марта 2018, Москва.

Публикации автора по теме диссертации. По теме диссертации опубликовано 15 работ в журналах и трудах конференций, 5 из них находятся в списке научных изданий, зарегистрированных в Высшей аттестационной комиссии Министерства науки и высшего образования России (в том числе 3 статьи в журналах из перечня RSCI – Web of Science).

Структура и объем диссертации. Диссертационная работа состоит из титульного листа, оглавления, введения, четырёх глав, заключения, списка литературы и 5-х приложений. Основная часть (без приложений) изложена на 154 страницах машинописного текса. Работа содержит 23 рисунка, 37 таблиц, список литературы включает 134 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы диссертации, сформулированы цель и задачи, объект и предмет исследования, научная новизна и практическая ценность диссертации, определена степень изученности темы, представлены основные научные положения, выносимые на защиту, приведен список Всероссийских и международных конференций с участием автора диссертации.

В первой главе «Аналитический обзор методов и алгоритмов в многоэкспертных системах принятия решений» описаны основные алгоритмы принятия решений МЭБС и известные приложения МЭБС. В заключении этой главы делаются выводы по материалам обзора и формулируются задачи, которые должны быть решены в диссертации.

Во второй главе «Структуры многоэкспертных бинарных систем и их оптимизация» содержится:

Анализ многоэкспертной бинарной системы принятия решения большинством голосов при равных значениях условных вероятностей принятия гипотезы каждого статистически взаимно независимого эксперта.

Пусть имеются n статистически независимых экспертов, на входы которых поступает сигнал x, принимающий два возможных значения, например, 0 и 1. Каждый i-й эксперт в соответствии с присущим ему статистическим критерием $f_i(x) = y_i$ проверяет статистические гипотезы H_0 и H_1, где гипотеза H_0 означает, что входной

сигнал имеет значение $x=0$, а гипотеза H_1 - значение $x=1$, и принимает решение о значении входного сигнала x. Точность принятия решения каждого i-го эксперта о значении входного сигнала характеризуется условными вероятностями: $p_i(0/0)$ - условная вероятность принять правильное решение о том, что входной сигнал равен 0 при условии, что значение входного сигнала равно 0, и $p_i(1/1)$- условная вероятность принять правильное решение о том, что входной сигнал равен 1, при условии, что значение входного сигнала равно 1. Величины $q_i(1/0)=1-p_i(0/0)$ и $q_i(0/1)=1-p_i(1/1)$ характеризуют условные вероятности неправильного принятия решения экспертом, а именно $q_i(1/0)$ - условная вероятность того, что эксперт i принял гипотезу H_1 (вероятность ошибки первого рода или вероятностью ложной тревоги), тогда как входной сигнал равнялся 0, а $q_i(0/1)$ - условная вероятность того, что эксперт i принял гипотезу H_0, тогда как входной сигнал равнялся 1 (вероятность ошибки второго рода или вероятностью пропуска события). Решения каждого эксперта $y_i, i=\overline{1,N}$ в виде N- мерного вектора Y поступают на верхний уровень МЭБС, в котором также имеется статистический критерий $f(Y)=z$, проверяющий статистические гипотезы H_0 и H_1. Все эксперты статистически независимы и обладают равной квалификацией, т.е. условные вероятности принять правильные решения и условные вероятности ошибок не зависят от номера эксперта:
$p_i(0/0)=p(0/0),\ q_i(1/0)=q(1/0),\ i=\overline{1,n}$
$p_i(1/1)=p(1/1),\ q_i(0/1)=q(0/1),\ i=\overline{1,n}$.

Вероятности принятия правильного решения в МЭБС при нечетном числе экспертов.

Обозначим $p_{expsys}(0/0),\ p_{expsys}(1/1), q_{expsys}(1/0),\ q_{expsys}(0/1)$ - условные вероятности правильных решений и вероятности ошибок МЭБС, тогда для рассматриваемого случая эти вероятности имеют вид:

$$p_{expsys}(0/0)=\sum_{k=0}^{(n-1)/2} C_N^k p^{n-k}(0/0)q^k(1/0), \quad (1)$$

$$p_{expsys}(1/1)=\sum_{k=0}^{(n-1)/2} C_n^k p^{n-k}(1/1)q^k(0/1), \quad (2)$$

$$q_{expsys}(1/0)=1-p_{expsys}(0/0),\ q_{expsys}(0/1)=1-p_{expsys}(1/1),$$

где C_n^k - число сочетаний из n по k: $C_n^k=\dfrac{n(n-1)...(n-k+1)}{k!}, C_n^0=C_n^n=1$.

На рисунке 1 представлены зависимости вероятностей правильного решения МЭБС (*ExpertSystPrb*) от условной вероятности правильного решения эксперта (*ExpertPrb*) при равной квалификации каждого эксперта для разных значений нечетного

числа статистически независимых экспертов, рассчитанные по формуле (1). Числа на кривых означают число экспертов системы. Из рисунка 1 следует:

1. Если условная вероятность правильного решения эксперта такой системы $p(0/0) > 0.5$, то найдется такое число экспертов, что вероятность $p_{expsys}(0/0)$ правильного решения МЭБС окажется как угодно близкой к 1, что подтверждает график функции, где число экспертов равно 10001. В этом случае функция практически имеет релейный характер.

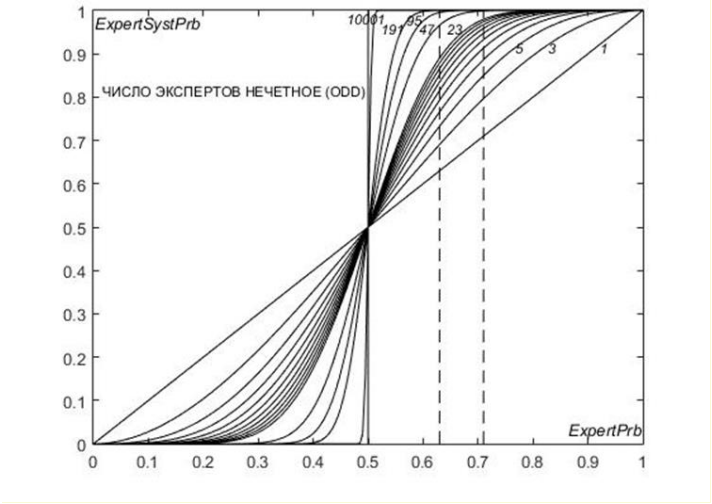

Рисунок 1

2. Если вероятность правильного решения эксперта $p(0/0) < 0.5$, то с увеличением числа экспертов результирующая вероятность $p_{expsys}(0/0)$ правильного решения МЭБС стремится к нулю.

3. Если вероятность правильного решения эксперта

$p(0/0) = 0.5$, то вероятность $p_{expsys}(0/0) = 0.5$ независимо от числа экспертов.

Вероятности принятия правильного решения в МЭБС при четном числе экспертов

Для четного числа экспертов формулы (1) и (2) могут быть записаны либо в виде

$$p_{expsys}(0/0) = \sum_{k=0}^{n/2} C_n^k p^{n-k}(0/0) q^k(1/0), \tag{3a}$$

$$p_{expsys}(1/1) = \sum_{k=0}^{n/2-1} C_N^k p^{n-k}(1/1) q^k(0/1), \tag{3b}$$

либо

$$p_{expsys}(0/0) = \sum_{k=0}^{n/2-1} C_N^k p^{n-k}(0/0) q^k(1/0), \tag{4a}$$

$$p_{expsys}(1/1) = \sum_{k=0}^{n/2} C_N^k p^{n-k}(1/1) q^k(0/1). \tag{4b}$$

На рисунке 2 и рисунке 3 представлены вероятности *ExpertSystPrb* правильного решения МЭБС в зависимости от условной вероятности правильного решения эксперта *ExpertPrb* при равной квалификации каждого эксперта для разных значений четного числа статистически независимых экспертов, рассчитанные по формулам (3a) и (3b), соответственно.

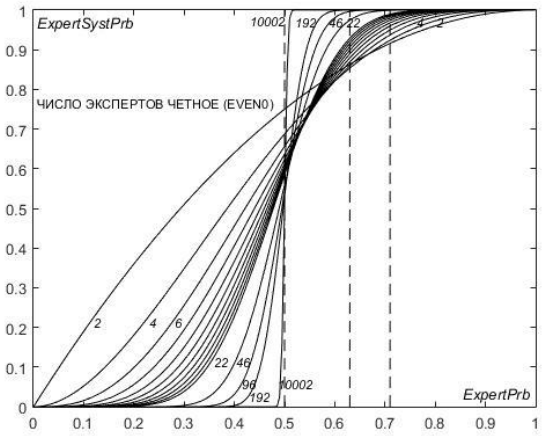

Рисунок 2

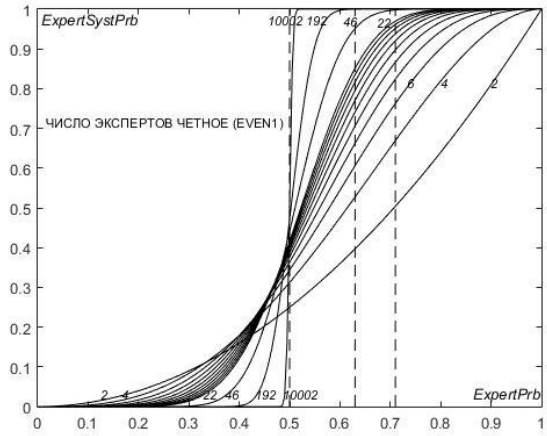

Рисунок 3

Из сравнения рисунка 2 и рисунка .3 следуют важные выводы и вопросы:

1. Влияние вероятности наступления элементарного события, когда гипотезы H_0 и H_1 принимают ровно половина экспертов, на вероятности правильного решения МЭБС $p_{expsys}(0/0)$ и $p_{expsys}(1/1)$ весьма велико. Им можно пренебречь при низкой квалификации экспертов только тогда, когда число экспертов более сотни. При больших значениях числа экспертов графики функций при четном и нечетном числе экспертов практически неотличимы. Наибольшая величина абсолютного отклонения функций $p_{expsys}(0/0)$ и $p_{expsys}(1/1)$ при четном числе экспертов $n=10002$, равная $\left|p_{expsys}(0/0)-p_{expsys}(1/1)\right|=0.00798$, достигается при $p(0/0)=p(1/1)=0.5$. Наибольшая величина абсолютного отклонения функций $p_{expsys}(0/0)$ и $p_{expsys}(1/1)$ при четном числе экспертов $n=10002$ от функции $p_{expsys}(0/0)$ при нечетном числе экспертов $n=10001$, равная 0.00399, также достигается при $p(0/0)=p(1/1)=0.5$.

2. Это влияние наиболее сильно прослеживается при небольшом числе экспертов.

Зависимость вероятности принятия правильного решения в МЭБС от нечетного и четного числа экспертов.

Формулы расчета вероятностей $p_{expsys}(0/0)$ (1) и $p_{expsys}(1/1)$ (2) принятия правильного решения в МЭБС при **нечетном** числе экспертов относительно гипотез H_0 и H_1 совпадают. Поэтому в этом случае расчет вероятностей правильного решения экспертной системы принятия решения большинством голосов в зависимости от числа

экспертов $n = 2m-1,\ m = 1,2,...$ выполняется только для вероятностей $p_{expsys}(0/0)$, которые на соответствующих рисунках обозначаются как $ExpertSystPrb$. Для наглядности последующего изложения вероятности $p_{expsys}(0/0)$, которые являются функциями числа экспертов n, будем обозначать как $p^{(0/0)}_{expsys}(n)$, а вероятности $p_{expsys}(1/1)$ - как $p^{(1/1)}_{expsys}(n)$.

Приведенные на рисунке 4 зависимости полностью подтверждают вывод в работе [1] о том, что при **нечетном** числе $n = 2m-1$ статистически независимых экспертов, для которых вероятность правильного решения равна p, имеют место следующее факты: 1) если $p > 0.5,$ то $p^{(0/0)}_{\exp sys}(2m-1)$ и $p^{(1/1)}_{\exp sys}(2m-1)$ - монотонно возрастающие функции m, при этом $p^{(0/0)}_{\exp sys}(2m-1)$ и $p^{(1/1)}_{expsys}(2m-1) \to 1,$ когда $m \to \infty$; 2) если $p < 0.5,$ то $p^{(0/0)}_{\exp sys}(2m-1)$ и $p^{(1/1)}_{\exp sys}(2m-1)$ - монотонно убывающие функции m, при этом $p^{(0/0)}_{\exp sys}(2m-1)$ и $p^{(1/1)}_{expsys}(2m-1) \to 1,$ когда $m \to \infty$; 3) если $p = 0.5$ то $p^{(0/0)}_{\exp sys}(2m-1)$ и $p^{(1/1)}_{\exp sys}(2m-1)$ равны 0.5 для всех m.

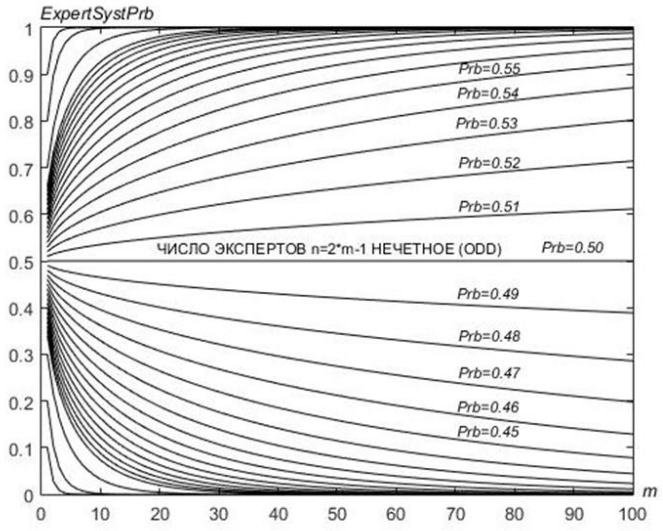

Рисунок 4

Из анализа формул (1) и (2), а также из рисунка 4 следует, что эти функции при вероятностях правильного решения статистически независимых экспертов, равных $Prb = 0.5 + \delta$ и $Prb = 0.5 - \delta,\ 0 < \delta < 0.5$, являются симметричными функциями относительно оси $ExpertSystPrb = 0.5$. Кроме того, чем меньше по абсолютному значению величина δ, тем медленнее стремятся соответствующие функции к 0 или 1.

Зависимости вероятности правильного решения экспертной системы принятия решения большинством голосов *ExpertSystPrb,* как функции **четного** числа экспертов $n = 2*m,\ m = 1,2,...$, существенно отличаются от аналогичных зависимостей при

нечетном числе экспертов, которые представлены на рисунке 4. Эти зависимости приведены на рисунке 5 и рисунке 6.

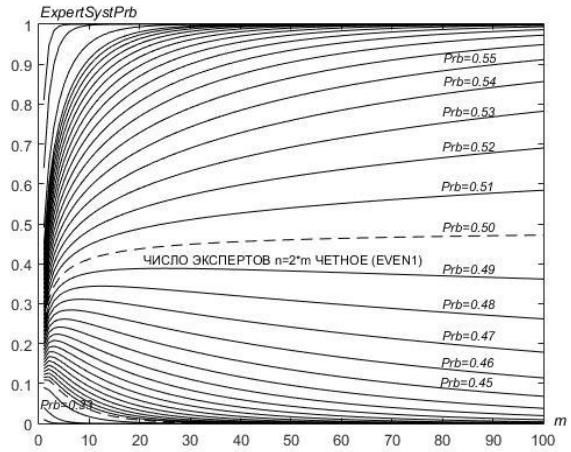

Рисунок 5

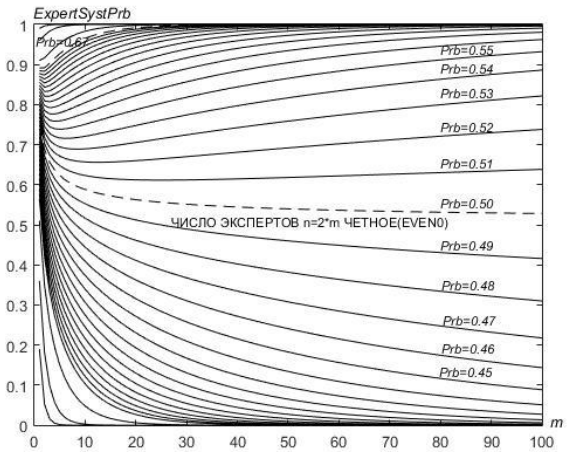

Рисунок 6

На рисунке 5 показаны зависимости $p_{expsys}^{(1/1)}(2m)$, рассчитанные по формуле (3b), которая соответствует вероятности правильного решения МЭБС относительно гипотезы H_1. Именно этот случай рассмотрен в работе [1], в которой доказано, что разность функций $p_{expsys}^{(1/1)}(2m)$ в соседних точках $m+1$ и m равна:

$$p_{expsys}^{(1/1)}(2(m+1)) - p_{expsys}^{(1/1)}(2m) = p^{m+1}q^n C_{2m}^m \left(\frac{2mp + p - m}{m+1} \right), \tag{5}$$

где $q = 1 - p$.

Анализ числителя скобки выражения (5) позволил авторам [1] сделать вывод о том, что при **четном** числе $n = 2m, m = 1,2,...$ статистически независимых экспертов, для которых вероятность правильного решения p удовлетворяет условиям: $0 < p < 1/3$ и $1/2 \le p < 1$, вероятности правильного решения $p_{expsys}^{(1/1)}(2m)$ многоэкспертной системы принятия решения большинством голосов являются монотонно убывающими и монотонно возрастающими функциями m, соответственно. В случае, когда $1/3 \le p < 1/2$, вероятности правильного решения $p_{expsys}^{(1/1)}(2m)$ многоэкспертной системы принятия решения большинством голосов не являются монотонными функциями числа экспертов системы и "поведение которых зависит от относительных величин p и $m/(2m+1)$".

Кроме того, рисунок 5, выражение (5) и анализ числовых значений функции $p_{expsys}^{(1/1)}(2m)$ позволяют сделать приведенные ниже новые заключения о характере функций $p_{expsys}^{(1/1)}(2m)$ и $p_{expsys}^{(0/0)}(2m)$ для гипотезы H_1.

Функции $p_{expsys}^{(1/1)}(2m)$, которым соответствуют значения вероятностей экспертов внутри области $1/3 < p < 1/2$, выделенные на рисунке 5 пунктирными линиями сначала возрастают, достигая максимума, не превышающего 0.5, после чего начинается их убывание. Функции имеют один максимум. Точка максимума определяется из условия равенства нулю числителя дроби

$$2mp + p - m \tag{6}$$

в выражении (5), откуда следует, что максимум функций $p_{expsys}(2m)$ достигается при значении

$$m = \beta = p/(1-2p). \tag{7}$$

Функции $p_{expsys}^{(1/1)}(2m)$ - функции целых значений аргумента m, поэтому если значение β в выражении (7) – целое число, то это означает, что (5) и (6) равны нулю, т.е. $p_{expsys}^{(1/1)}(2(m+1)) = p_{expsys}^{(1/1)}(2m)$, и в этом случае максимум достигается в точках m и $m+1$. Если же β в выражении (7) нецелое число, то нетрудно показать, что максимум достигается в точке $m = \mathrm{int}(\beta+1)$, где функция int - целая часть своего аргумента. Окончательно, функции $p_{expsys}^{(1/1)}(2m)$ достигают своих максимальных значений в точках m^*, где:

$$m^* = \begin{cases} m,\ m+1,\ \textit{где}\ m = \beta = p/(1-2p), & \textit{если}\ \beta\ \textit{целое}, \\ \mathrm{int}(\beta+1), & \textit{если}\ \beta\ \textit{нецелое}. \end{cases} \tag{8}$$

Выражение (6) позволяет определить те значения вероятностей p^*, при которых максимум функций $p_{expsys}^{(1/1)}(2m)$ достигается в точках m и $m+1$:

$$p^* = m/(1+2m),\ m = 1, 2, \ldots. \tag{9}$$

Зависимость вероятности принятия правильного решения в МЭБС при четном числе экспертов в случае гипотезы H_0

Вероятности $p_{expsys}^{(0/0)}(2m)$ относительно гипотезы H_0 рассчитываются по формуле (3а), которая при $n = 2m$ принимает вид:

$$p_{expsys}^{(0/0)}(2m) = \sum_{k=0}^{m} C_{2m}^{k} p^{2m-k} q^{k}\ . \tag{10}$$

После довольно громоздких преобразований, приведенных в диссертации, с учетом соотношения для сочетаний $C_n^m = C_{n-1}^m + C_{n-1}^{m-1}$, получим соотношение для первой разности при четном числе экспертов:

$$p_{expsys}^{(0/0)}(2(m+1)) - p_{expsys}^{(0/0)}(2m) = C_{2m+1}^m p^m q^{m+1}\left(p - \frac{m+1}{2m+1}\right). \quad (11)$$

Соотношение (11) применительно к гипотезе H_0 является аналогом соотношения (5) применительно к гипотезе H_1. Это соотношение позволяет провести анализ поведения функций $p_{expsys}^{(0/0)}(2m)$ в зависимости от четного числа экспертов. Из (11) следует, что поведение функции $p_{expsys}^{(0/0)}(2m)$ определяется знаком выражения

$$\psi(p,m) = p - \frac{m+1}{2m+1} = p - \varphi(m). \quad (12)$$

Поскольку функция $\varphi(m)$ удовлетворяет условию: $1/2 < \varphi(m) \le 2/3$, то для всех $2/3 < p < 1$ функция $p_{expsys}^{(0/0)}(2m)$- монотонно возрастающая функция m, для всех $0 < p \le 1/2$ - монотонно убывающая функция m. Отметим, что при $0 < p < 1/2$ пределом функций $p_{expsys}^{(0/0)}(2m)$ является 0, а при $p = 1/2$ пределом служит 0.5.

Учитывая, что функция $\varphi(m)$ - монотонно убывающая функция m, то для значений p, удовлетворяющих условию $1/2 < p < 2/3$, функция $p_{expsys}^{(0/0)}(2m)$ сначала убывает с ростом m, достигая своего минимального значения, после чего начинается ее рост. Рисунок 2.7 полностью подтверждает эти выводы. Кроме того, функция $p_{expsys}^{(0/0)}(2m)$ при $p = 2/3$ принимает минимальные значения при $m = 1$ и $m = 2$.

Минимум функции $p_{expsys}^{(0/0)}(2m)$ достигается, когда $\psi(p,m) = 0$, т.е. при

$m = \frac{1-p}{2p-1}$. Поскольку m - целые числа, выражение для m по аналогии с (8) имеет вид

$$m^* = \begin{cases} m,\ m+1,\ \textit{где}\ m = \gamma = (1-p)/(2p-1), & \textit{если}\ \beta\ \textit{целое}, \\ \\ \text{int}(\gamma+1), & \textit{если}\ \gamma\ \textit{нецелое}. \end{cases} \quad (13)$$

Еще несколько выводов относительно функции $p_{expsys}^{(0/0)}(2m)$ при значениях p, удовлетворяющих условию $1/2 < p < 2/3$. Чем ближе значение p к $1/2$, тем при больших значениях m достигается минимальное значение и тем меньше значение минимума, не превышающего значения 0.5.

Оптимальная многоэкспертная бинарная система голосования при равных значениях условных вероятностей принятия гипотезы H_0 и H_1 каждого статистически взаимно независимого эксперта.

В настоящей работе в качестве критерия оптимальности МЭБС принята линейная комбинация вероятностей ее правильного решения:

$$J(\alpha) = \alpha\, p_{expsys}(0/0) + (1-\alpha)\, p_{expsys}(1/1),\ \ 0 < \alpha < 1, \tag{14}$$

где α - весовой коэффициент, характеризующий предпочтение относительно принятия гипотез правильного решения системы. Задача синтеза оптимальной системы - нахождение таких параметров системы, при которых функционал (14) достигает наибольшего значения.

В общем случае, когда система состоит из N экспертов, можно говорить о различных процедурах голосования, когда гипотезу H_0 принимают более $(n-1)$ экспертов, $n = \overline{1,N}$, а гипотезу H_1 принимают более $(N-n)$ экспертов. В этом случае, как вероятности правильного решения системы

$$p^{(n)}_{expsys}(0/0) = \sum_{k=0}^{N-n} C_N^k p^{N-k}(0/0) q^k(1/0),$$
$$p^{(N-n+1)}_{expsys}(1/1) = \sum_{k=0}^{n-1} C_N^k p^{N-k}(1/1) q^k(0/1),\ n = \overline{1,N}, \tag{15}$$

так и функционал (14), характеризующий свойство МЭБС, оказываются функциями числа n. Значение числа $n = n_{opt}$, при котором функционал (14) достигает своего максимального значения, определяет оптимальную процедуру голосования:

$$n_{opt} = \arg\max_n (\alpha\, p^{(n)}_{expsys}(0/0) + (1-\alpha)\, p^{(N-n+1)}_{expsys}(1/1),\ \ 0 < \alpha < 1), n = \overline{1,N}\,. \tag{16}$$

В качестве примера в таблице 1 показаны результаты 7 вариантов различных процедур голосования в системе, состоящей из 7 независимых экспертов. Здесь приведены условные вероятности правильного принятия решений системой (2 и 3 строки) и соответствующие значения функционала при трех значениях параметра α (4, 5 и 6 строки таблицы). Оптимальные значения функционала выделены серым цветом.

Таблица 1. Значения условных вероятностей $p^{(n)}_{expsys}(0/0)$, $p^{(N-n+1)}_{expsys}(1/1)$ правильного решения МЭБС и соответствующих функционалов для $n = \overline{1,7}$ при условных вероятностях правильного решения статистически независимых 7 экспертов $p_i(0/0) = 0.98$ и $p_i(1/1) = 0.6,\ i = \overline{1,7}$.

	$n = 7$	$n = 6$	$n = 5$	$n = 4$	$n = 3$	$n = 2$	$n = 1$
$p^{(n)}_{expsys}(0/0)$	0.86813	0.99214	0.99974	0.99999	1.00000	1.00000	1.00000

$p_{expsys}^{(N-n+1)}(1/1)$	0.99836	0.98116	0.90374	0.71021	0.41990	0.15863	0.02799
$J(\alpha = 0.5)$	0,93324	0,98665	0,95174	0,85510	0,70995	0,57932	0,51400
$J(\alpha = 0.9)$	0,88115	0,99104	0,99014	0,97101	0,94199	0.91586	0.90280
$J(\alpha = 0.1)$	0,98534	0,98226	0,91334	0,73919	0,47791	0,24277	0,12519

Оптимальная МЭБС голосования при неравных значениях условных вероятностей принятия гипотез H_0 и H_1 каждого статистически взаимно независимого эксперта

Рассмотрим общий случай при неравных значениях условных вероятностей, когда $p_i(0/0) \neq p(0/0);\ p_i(1/1) \neq p(1/1),\ i = \overline{1,N}$.

Можно показать, что формулы (15) в этом случае принимают вид:

$$p_{expsys}^{(n)}(0/0) = \prod_{i=1}^{N} p_i(0/0) + \sum^{C_N^1} p_{i_1}(0/0)\, p_{i_2}(0/0)...p_{i_{N-1}}(0/0)q_{i_N}(1/0) + ... + \sum^{C_N^{N-n}} p_{i_1}(0/0)\, p_{i_2}(0/0)...p_{i_{N-n}}(0/0)q_{i_{N-n+1}}(1/0)q_{i_{N-n+2}}(1/0)...q_{i_N}(1/0), \quad (17)$$

$$p_{expsys}^{(N-n+1)}(1/1) = \prod_{i=1}^{N} p_i(1/1) + \sum^{C_N^1} p_{i_1}(1/1)\, p_{i_2}(1/1)...p_{i_{N-1}}(1/1)q_{i_N}(0/1) + ... + \sum^{C_N^{n-1}} p_{i_1}(1/1)\, p_{i_2}(1/1)...p_{i_{n-1}}(1/1)q_{i_n}(1/0)q_{i_{n+1}}(0/1)...q_{i_N}(0/1). \quad (18)$$

В формулах (17), (18) значения индексов $i_j,\ j = \overline{1,N}$ не совпадают.

Ниже приведены результаты расчета оптимальной МЭБС голосования при 7 экспертах, вероятности правильного принятия гипотез которых $p_i(0/0), p_i(1/1),\ i = \overline{1,7}$ представлены в таблице 2.

Таблица 2. Значения условных вероятностей правильного решения 7-и экспертов

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$
$p_i(0/0)$	0.99	0.98	0.97	0.96	0.95	0.94	0.93
$p_i(1/1)$	0.69	0.68	0.67	0.66	0.65	0.64	0.63

Результаты расчета условных вероятностей правильного принятия гипотез системой для экспертов разной квалификации (таблица 2) по формулам (17), (18) и значения соответствующих функционалов приведены в таблице 3.

Таблица 3. Значения условных вероятностей правильного решения системы для экспертов разной квалификации (таблица 2) и значения соответствующих функционалов

	$n = 7$	$n = 6$	$n = 5$	$n = 4$	$n = 3$	$n = 2$	$n = 1$
$p_{expsys}^{(n)}(0/0)$	0.75031	0.97152	0.99824	0.99994	1.00000	1.00000	1.00000
$p_{expsys}^{(N-n+1)}(1/1)$	0.99948	0.99239	0.95095	0.81653	0.55527	0.25099	0.05438
$J(\alpha = 0.5)$	0.87490	0.98196	0.97460	0.90824	0.77764	0.62550	0.52719
$J(\alpha = 0.9)$	0.77523	0.97361	0.99351	0.98160	0.95553	0.92510	0.90544
$J(\alpha = 0.1)$	0.97456	0.99030	0.95568	0.83487	0.59974	0.32589	0.14894

Оптимальная МЭБС голосования при неравных значениях условных вероятностей принятия гипотез H_0 и H_1 каждого статистически взаимно зависимого эксперта на основе метода статистических испытаний

Задача существенно усложняется, когда эксперты являются статистически зависимыми. В этом случае, если даже удается построить совместные плотности распределений правильных решений экспертов, то применение их для расчета условных вероятностей типа (17), (18) вряд ли даст аналитические решения.

Подход к решению рассматриваемой задачи на основе метода статистических испытаний заключается в следующем. На вход МЭБС или ее модели подается случайная последовательность $X(m)$, m -номер шага последовательности. Входная последовательность, состоящая из нулей и единиц, поступает на входы всех экспертов системы, рисунок 7. Значения вероятностей правильных решений экспертов априори неизвестны. Неизвестно также, имеются ли статистические связи между отдельными экспертами.

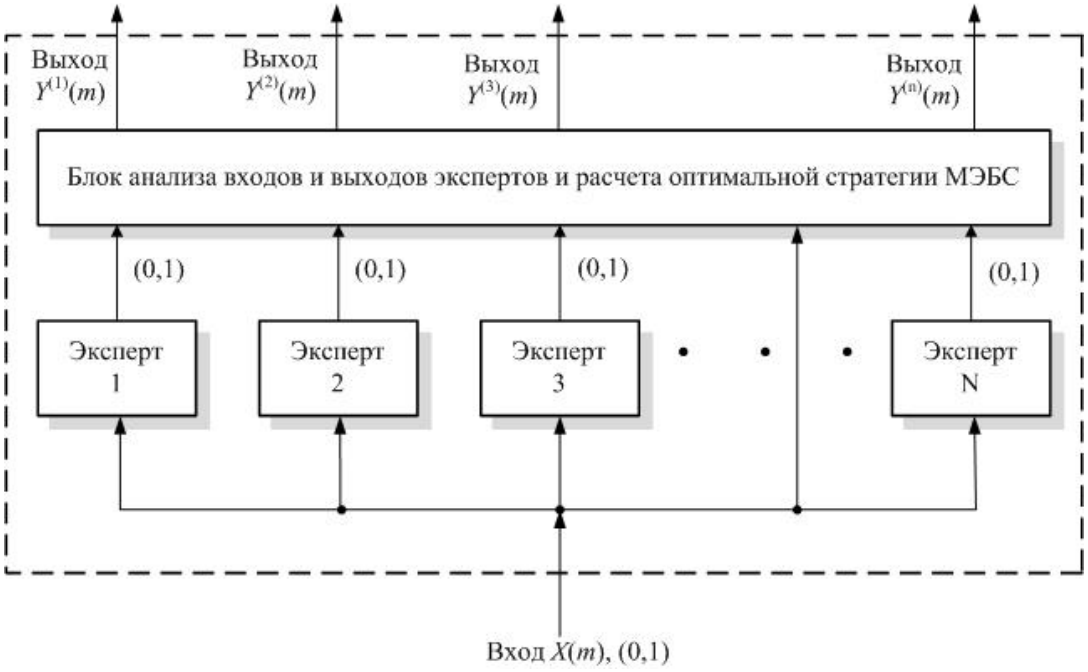

Рисунок 7. Структурная схема МЭБС в режиме статистических испытаний

Значения входа в МЭБС и выходов экспертов поступают в блок анализа, который на основании этих данных вычисляет оценки условных вероятностей правильных решений экспертов $\hat{p}_i(0/0), \hat{p}_i(1/1), i=\overline{1,N}$ и оценки условных вероятностей того, что гипотезу H_0 или H_1 принимают более $(n-1)$-го эксперта: $\hat{p}^{(n)}_{expsys}(0/0)$ и $\hat{p}^{(n)}_{expsys}(1/1)$. В блоке анализа МЭБС заложено значение параметра α функционала (14) для нахождения номера оптимального выхода $Y^{(n)}(m)$:

$$n_{opt}=\arg\max_n(\alpha\,\hat{p}^{(n)}_{expsys}(0/0)+(1-\alpha)\hat{p}^{(N-n+1)}_{expsys}(1/1),\ \ 0<\alpha<1), n=\overline{1,N}. \qquad (19)$$

Для вычисления оценок вероятностей правильных решений экспертов вводятся переменные $M0(m), M1(m)$ - число нулей и единиц во входной последовательности длиной m, соответственно, и $M0_i(m), M1_i(m), i=\overline{1,N}$ - число правильных решений экспертов относительно гипотез H_0 или H_1, принятых экспертами. Если на следующем $(m+1)$-ом шаге испытаний входная последовательность принимает значение $X(m+1)=0$, то переменная $M0(m)$ увеличивается на единицу и на единицу увеличиваются те значения переменных $M0_i(m)$, i-е эксперты которых дали правильное решение, равное 0. Аналогичная процедура выполняется, если входная переменная принимает значение $X(m+1)=1$. Оценки условных вероятностей правильных решений экспертов вычисляются по формулам

$$\hat{p}_i(0/0)=M0_i(m)/M0(m),\ \ \hat{p}_i(1/1)=M1_i(m)/M1(m),\ i=\overline{1,N}. \qquad (20)$$

Оценки (20) могут служить для оптимизации МЭБС, связанной с исключением экспертов низкой квалификации из состава МБЭС.

Для расчета условных вероятностей $\hat{p}^{(n)}_{expsys}(0/0)$ и $\hat{p}^{(n)}_{expsys}(1/1)$, необходимых для нахождения оптимальной МЭБС, вводятся переменные $M0^{(n)}_{\exp sys}(m), M1^{(n)}_{expsys}(m), n=\overline{1,N}$, пересчет которых ведется по следующему правилу: если на следующем $(m+1)$-ом шаге испытаний входная последовательность принимает значение $X(m+1)=0$, то определяется число экспертов $L0$, которые дали правильное решение, равное 0. Далее часть значений переменных $M0^{(n)}_{expsys}(m)$ с номерами $n=\overline{1,L0}$ увеличивается на единицу: $M0^{(n)}_{expsys}(m)=M0^{(n)}_{expsys}(m)+1, n=\overline{1,L0}$. Значения всех переменных $M1^{(n)}_{expsys}(m), n=\overline{1,N}$ в этом случае остаются без изменения. Если входная переменная принимает значение $X(m+1)=1$, то определяется число экспертов $L1$, которые дали правильное решение, равное 1. Часть значений переменных $M1^{(n)}_{expsys}(m)$ увеличивается на единицу: $M1^{(n)}_{expsys}(m)=M1^{(n)}_{expsys}(m)+1, n=\overline{1,L1}$. Значения всех переменных в этом случае остаются без изменения. $M0^{(n)}_{expsys}(m), n=\overline{1,N}$

Оценки условных вероятностей $\hat{p}^{(n)}_{expsys}(0/0)$ и $\hat{p}^{(n)}_{expsys}(1/1)$ того, что гипотезу H_0 или H_1 принимают более $(n\text{-}1)$ экспертов, теперь вычисляются по формулам

$$\begin{aligned}&\hat{p}^{(n)}_{expsys}(0/0)=M0^{(n)}_{expsys}(m)/M0(m),\\&\hat{p}^{(n)}_{expsys}(1/1)=M1^{(n)}_{expsys}(m)/M1(m),\ n=\overline{1,N}\end{aligned}, \qquad (21)$$

которые являются основой для синтеза оптимальной МЭБС. Точность оценок (21), и, следовательно, точность нахождения оптимальной МЭБС в режиме статистических испытаний существенно зависит от числа шагов входной последовательности $X(m)$ и от отношения нулей и единиц в этой последовательности.

В диссертации приведены результаты программной реализации режима статистических испытаний для оптимизации МЭБС, которые показали эффективность данного подхода. Показано, что решение системы в режиме статистических испытаний асимптотически приближается к точному решению, когда известны значения условных вероятностей экспертов.

О составе экспертов в МЭБС

В МЭБС, которая состоит из экспертов разной квалификации, могут присутствовать нежелательные эксперты, например эксперты, вероятности принятия правильных решений $p_i(0/0)$ и $p_i(1/1)$ которых меньше 0.5. Удаление таких экспертов из состава экспертов МЭБС приведет к увеличению значения критерия (14).

При формальном подходе анализа состава экспертов следует поочередно исключать экспертов из состава, проводить оптимизацию системы с уменьшенным числом экспертов, после чего удалять того эксперта, исключение которого привело к наибольшему увеличению значения критерия. Далее подобная процедура повторяется к системе с уменьшенным числом экспертов до тех пор, пока не прекратиться увеличение значения критерия. Такой подход гарантирует правильный конечный результат, но является достаточно ресурсоемким.

Более эффективным служит подход на основе ранжирования, который основывается на том, что каждому эксперту с номером i ставится в соответствие критерий, по структуре и параметрам совпадающий со структурой критерия (14) МЭБС:

$$J_i(\alpha) = \alpha\, p_i(0/0) + (1-\alpha)\, p_i(1/1),\ \ 0<\alpha<1, i=\overline{1,N}, \tag{22}$$

где $p_i(0/0)$ и $p_i(1/1)$ - вероятности правильных решений i-го эксперта относительно гипотез H_0 и H_1. Далее определяется номер эксперта $i^* = \min\limits_i J_i(\alpha)$, и данный эксперт удаляется из состава экспертов. Для уменьшенного состава экспертов находится оптимальное значение критерия (22).

В третьей главе «**Нейросетевые технологии обнаружение атак**» основное внимание уделяется проблеме обучения отдельных экспертов системы обнаружения атак на основе нейронной сети CMAC, многослойных нейронных сетей (Multilayer Neural Network), метода случайного леса (Random Forest), метода опорных векторов (Support Vector Machines). Материалы этой главы служат подготовительным этапом для создания оптимальной многоэкспертной системы.

Проведен анализ используемых баз данных атак KDD Cup 99 и UNSW-NB 15, которые необходимы для обучения отдельных экспертов. Предложен и исследован подход, основанный на комбинации метода случайного леса и многослойной нейронной сети для нахождения вектора признаков наименьшей размерности, который необходим для обучения нейросетевых экспертов. Особое внимание уделено обучению и применению НС CMAC для обнаружения DoS-атак на основе информации из базы данных UNSW-NB. Процесс использования НС CMAC в данной задаче представлен на рисунке 8.

Рисунок 8 - Иллюстрация процесса применения НС CMAC

В четвертой главе «**Многоэкспертные бинарные системы как средство повышения вероятности обнаружения атак**» приведены состав и процесс построения программ решений МЭБС для обнаружения атак типа Reconnaissance и DoS.

Многоэкспертная бинарная система для обнаружения Reconnaissance атак (R-атак) в базе данных UNSW-NB15.

Для создания МЭБС в задаче обнаружения атак в качестве экспертов использованы эксперты на основе многослойных нейронных сетей (МНС), метода опорных векторов (SVM – support vector machine) и метода случайного леса (RF – random forest).

Структура многоэкспертной бинарной системы, состоящей из пяти бинарных экспертов, представлена на рисунке 9. Структура многоэкспертной бинарной системы, состоящей из трех экспертов, представлена только в виде 3-х МНС.

Структура многоэкспертной бинарной системы, состоящей из пяти экспертов:

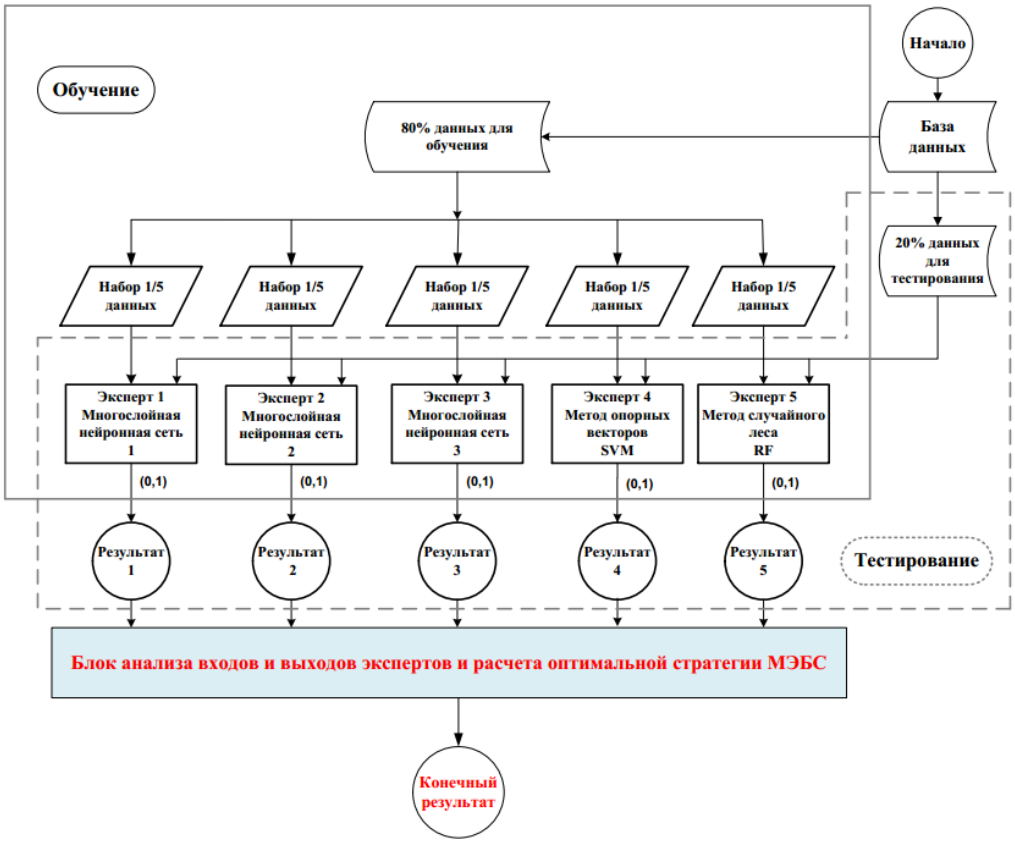

Рисунок 9. Структура МЭБС, состоящей из пяти экспертов

Оптимизация МЭБС, состоящей из трех экспертов

В случае применения алгоритма принятия решения большинством голосов система сообщает об обнаружении R-атаки, если два или три эксперта сообщают, что происходит атака. Если же два или три эксперта сообщают, что имеет место не R-атака, то система сообщает, что R-атака отсутствует - имеет место не R-атака. При таком подходе распознается 85.71 % атак типа Reconnaissance и 95.06% случаев остальных соединений.

Оптимизация МЭБС голосования выполнена путем изменения принципа голосования следующим образом: если три эксперта голосуют, что имеет место не R-атака, то система сообщает об отсутствии R-атаки, в остальных случаях система сообщает об обнаружении R-атаки. Результат для такой системы: распознается 96.23% атак R и 91.07% случаев остальных соединений.

Кроме этих двух возможных решений МЭБС, существует и третье, когда атака принимается, если за нее голосуют три эксперта. Все три возможных решения для данной системы представлены в таблице 4.

Таблица 4. Решения МЭБС, состоящей из 3-х экспертов

№ решения	Число R-атак в	Число не R-атак	Число не R атак,	Число обнаруж-енных	Процент обнаруж-енных	Процент обнаружен-ных не

		выборке	в выборке	принятых за R-атаку	R-атак	R-атак , %	R-атак, %
1	3 эксперта сообщают о не R-атаке	2043	30727	2744	1966	96.23	91.07
2	2 или 3 эксперта сообщают о не R-атаке	2043	30727	1519	1751	85.71	95.06
3	1, 2 или 3 эксперта сообщают о не R-атаке	2043	30727	459	1384	67.74	98.51

Оптимизация МЭБС, состоящей из пяти экспертов.

Таблица 5. 5 решений МЭБС, состоящей из пяти экспертов.

№ решения		Число R- атак в выборке	Число не R-атак в выборке	Число не R-атак, Принятых за R- атаку	Число обнаруженных R-атак	Процент обнаруженных R-атак, %	Процент обнаруженных не R-атак, %
1	5 экспертов сообщают о не R-атаке	2043	30727	3868	2018	98.78	87.41
2	4 или 5 экспертов сообщают о не R-атаке	2043	30727	2699	1983	97.06	91.22
3	3, 4 или 5 экспертов сообщают о не R атаке	2043	30727	2195	1871	91.58	92.86
4	2,3,4 или 5 экспертов сообщают о не R атаке	2043	30727	1497	1629	79.74	95.13
5	1,2,3,4 или 5 экспертов	2043	30727	490	1301	63.68	98.41

	сообщают о не R атаке						

Из анализа таблицы 5 следует, что оптимальная МЭБС с пятью экспертами предоставляет возможность в выборе того варианта решения, который позволит решить задачу обнаружения R-атаки или не R-атаки в соответствии с функционалом, который заложен в систему. В отличие от системы большинство голосов, представленной третьей строкой таблицы 4.6, с вероятностью обнаружения R-атаки, равной 0.92, и вероятностью обнаружения не R-атаки, равной 0.93, данная система предоставляет еще 4 варианта решений, которые могут быть использованы согласно заложенному в систему функционалу.

Сравнение результатов после оптимизации МЭБС, состоящей из трех и пяти экспертов, представлено на рисунке 10.

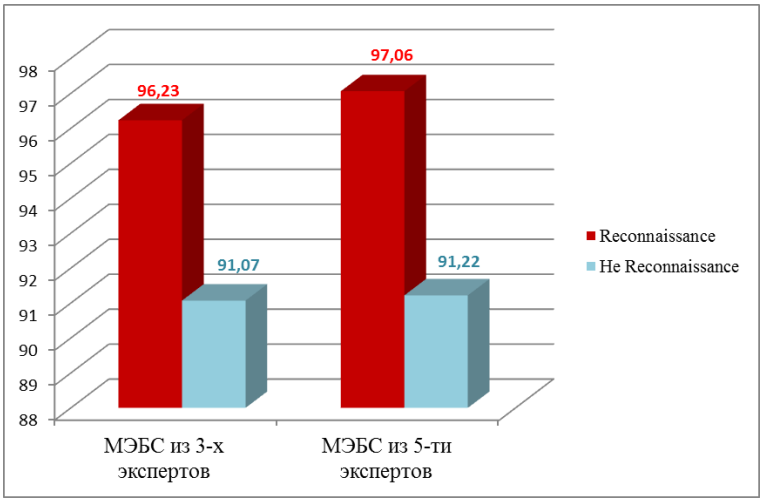

Рисунок 10. Сравнение результатов многоэкспертных бинарных систем, состоящих из трех и пяти экспертов

Многоэкспертные бинарные системы обнаружения атак типа DoS в базе данных UNSW-NB 15

Структура многоэкспертной бинарной системы, состоящей из шести экспертов

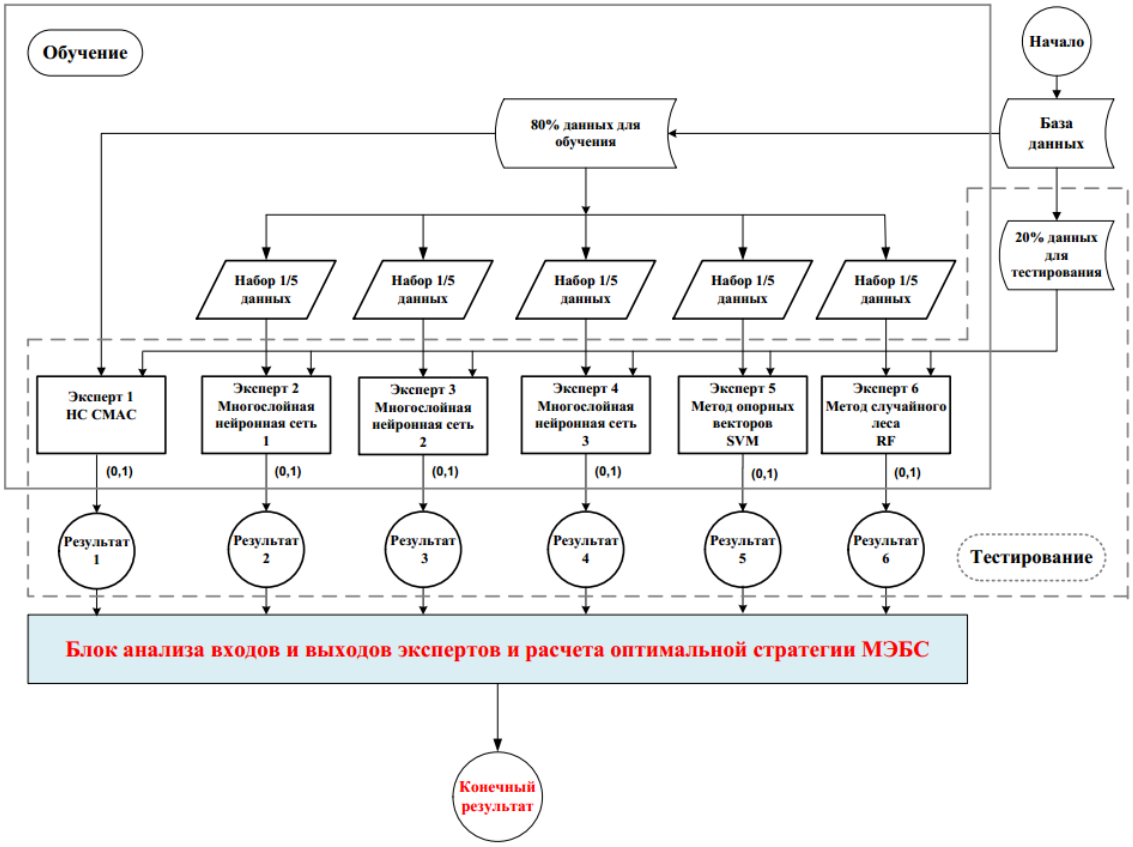

Рисунок 11. Структура МЭБС, состоящей из шести экспертов

Для трех экспертов в виде МНС были проведено обучение и тестирование трехслойных (15-10-1, 30-20-1, 50-30-1, 100-50-1, 100-100-1, 150-100-1, 200-100-1, 200-150-1) и четырехслойной нейронной сети (30-20-10-1). Были использованы разные методы обучения: метод Левенберга-Марквардта (trainlm), метод градиентного спуска с учетом моментов и с адаптивным обучением (traingdx), метод градиентного спуска с учетом моментов (traingdm), метод шкалированных сопряженных градиентов (trainscg), метод градиентного спуска (traingd), метод градиентного спуска с адаптивным обучением (traingda), алгоритм упругого обратного распространения (trainrp), квази-Ньютононовский метод, использующий BFGS (trainbfg). Эксперт по методу SVM использовал два разных значения функций ядра: Gaussian Radial Basis Function (RBF) и polynomial. Эксперт по методу случайного леса, который базируется на ансамбле из большого числа решающих деревьев, использовал подвыборку размером N, где N= 42 - число информативных признаков сетевого соединения.

Процесс оптимизация МЭБС, состоящей из шести экспертов:

Решения МЭБС_6 приведены в таблице 6.

Таблица 6. Решении МЭБС_6

№ решения	Число DoS-атак в выборке	Число не DoS-атак в выборке	Число не DoS-атак, принятых за DoS-атак	Число обнаруженных DoS-атак	Процент обнаруженных DoS-атак, %	Процент обнаруженных не DoS-атак, %

1	6 экспертов сообщают о DoS-атаке	1103	31616	294	79	7.16	99.07
2	5 или 6 экспертов сообщают о DoS-атаке	1103	31616	1036	309	28.01	96.72
3	4, 5 или 6 экспертов сообщают о DoS-атаке	1103	31616	2015	564	**51.13**	**93.62**
4	3,4,5 или 6 экспертов сообщают о DoS-атаке	1103	31616	3022	795	72.08	90.43
5	2,3,4,5 или 6 экспертов сообщают о DoS-атаке	1103	31616	4673	965	87.49	85.20
6	1,2,3,4,5 или 6 экспертов сообщают о DoS-атаке	1103	31616	7357	1067	96.34	76.71

Из анализа таблицы 6 следует что оптимальная МЭБС с шестью экспертами предоставляет возможность в выборе того варианта решения, который позволит решить задачу обнаружения DoS-атаки или не DoS-атаки в соответствии с функционалом, который заложен в систему. В отличие от системы принятия решения большинством голосов, представленной третьей строкой таблицы 6, с вероятностью обнаружения атаки DoS-атаки, равной 0.5113, и вероятностью обнаружения не DoS-атаки, равной 0.9362, данная система предоставляет еще 5 вариантов решений.

Иллюстрация программной реализации решения МЭБС

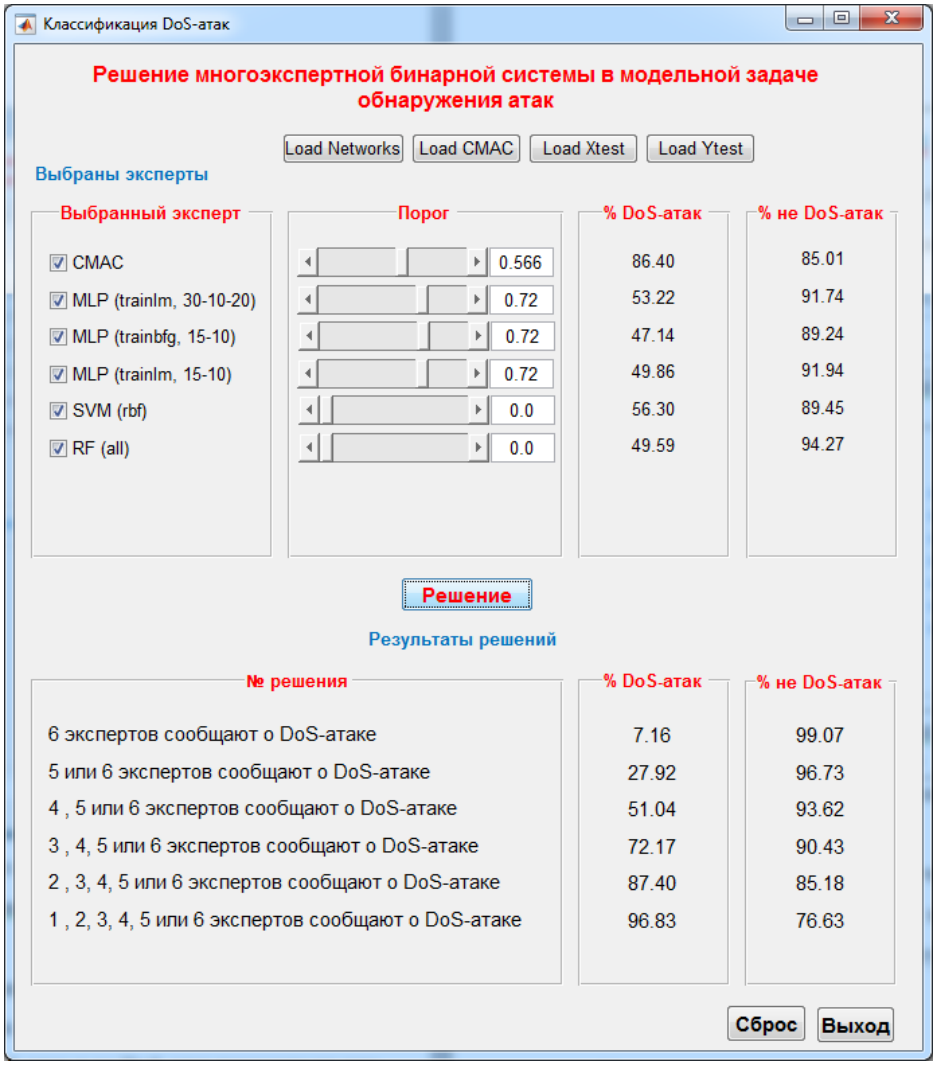

Рисунок 12. Иллюстрация интерфейса программной реализации решений МЭБС

В **Заключении** подведены основные результаты диссертационного исследования.

Иллюстративная 4-я глава диссертационной работы показывает, как подход к оптимизации многоэкспертных бинарных систем, описанный и исследованный во 2-й главе, может быть применен в задачах оптимизации решений систем поддержки принятия решений применительно к системам обнаружения атак.

В ходе исследований по оптимизации многоэкспертных систем, основанных на принципах голосования, и их применения в модельной системе обнаружения атак получены следующие основные результаты:

1. Установлены новые свойства функций $p_{expsys}^{(1/1)}(2m)$ четного числа экспертов, характеризующих вероятности принятия правильного решения МЭБС, когда за гипотезу H_1 голосует более половины экспертов при равной вероятности правильного решения каждого эксперта p. Показано, что данные функции являются немонотонными при $1/3 < p < 1/2$ и одноэкстремальными. Определены точки максимума функций. Эти результаты расширяют факты, приведенные в работе [1].

2. Установлены свойства функций $p_{expsys}^{(0/0)}(2m)$ четного числа экспертов, характеризующих вероятности принятия правильного решения МЭБС, когда за альтернативную гипотезу H_0 голосует или половина или более половины экспертов при равной вероятности правильного решения каждого эксперта p. Показано, что с ростом числа экспертов при значениях вероятности p, удовлетворяющих условию $2/3 < p < 1$, функции $p_{expsys}^{(0/0)}(2m)$ являются монотонно возрастающими и стремящимися к 1, при $0 < p < 1/2$ - эти функции - монотонно убывающие, стремящиеся к нулю. При p, удовлетворяющих условию $1/2 < p < 2/3$, функции $p_{expsys}^{(0/0)}(2m)$ сначала с увеличением числа экспертов убывают, достигают минимума, после чего начинается их рост и стремление к 1. Определены точки минимума функций.

3. Исследование МЭБС принятия решения большинством голосов показало, что принцип голосования в системах такого рода не является наилучшим.

4. Сформулирован принцип построения оптимальных МЭБС, основанных на принципах голосования, при котором определяется число экспертов, голосующих за каждую из двух гипотез так, чтобы функционал, представляющий линейную комбинацию вероятностей правильного решения МЭБС относительно каждой из гипотез, достигал максимального значения.

5. Этот принцип исследован в случае, когда известны значения условных вероятностей экспертов системы и имеет место их взаимная статистическая независимость.

6. Изложен подход для случая, когда условные вероятности экспертов системы неизвестны и может иметь место их статистическая зависимость. Данный подход

базируется на методе статистических испытаний на модели МЭБС или непосредственно на ней самой.

7. Проведенный глубокий экспериментальный анализ показал, что НС CMAC обладает большой перспективностью как аналитический инструмент обнаружения атак с высокой точностью, но с ограниченным количеством входных признаков.

8. На примере комплекса нейронных сетей CMAC, обученных обнаружению только одного типа атак, представлен подход обнаружения целого комплекса типов атак.

9. Изложен подход к синтезу оптимальной МЭБС, предназначенной для обнаружения атак, которая состоит из экспертов различной структуры и природы.

10. Приведены результаты компьютерного моделирования, которые иллюстрируют и подтверждают теоретические выводы, полученные в работе, а также результаты компьютерного моделирования оптимальных МЭБС, предназначенных для обнаружения атак, которые характеризуют эффективность таких систем.

Приложения содержатся таблицы признаков баз данных KDD Cup 99 и UNSW NB 15, коды для компьютерного моделирования решений МЭБС, акт о внедрении результатов разработки.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

В изданиях, входящих в перечень RSCI:

1. Аведьян Э. Д., Ле Т. Ч. Л. Двухуровневая система обнаружения DoS-атак и их компонентов на основе нейронных сетей CMAC // Информационные технологии. Том 29. 2016. № 9. С. 711-718.
2. Аведьян Э. Д., Ле Тхи Чанг Линь. Процедуры оптимального голосования в многоэкспертных бинарных системах. Труды МФТИ. Том 9, № 4 (36). 2017. С. 174-189 .
3. Ле Тхи Чанг Линь. Многоэкспертные бинарные системы как средство повышения вероятностей обнаружения атак на информационные ресурсы. Труды МФТИ. Том 10, № 1 (37). 2018. С. 155-167 .

В изданиях, входящих в перечень ВАК при Министерстве науки и высшего образования России:

1. Аведьян Э. Д., Ле Т. Ч. Л. Нейронная сеть CMAC в задаче обнаружения атак на информационные ресурсы // Информатизация и связь. 2015. № 4. С. 93 – 98.
2. Т. Ч. Л. Ле. Сравнение нейронной сети CMAC и многослойной нейронной сети в задаче обнаружения DoS-атак // Нейрокомпьютеры: разработка, применение. 2016. № 7. С. 65 – 69.
3. Ле Т.Ч.Л. Обнаружение атак в современной базе данных UNSW-NB15 с применением многослойной нейронной сети// Информатизация и связь. 2017. № 1. С. 61 – 66.

В других изданиях:

1. Аведьян Э.Д., Ле Т. Ч. Л. Технология обнаружения атак на основе нейронной сети CMAC//Сборник научных трудов по материалам IX Международной научно - практической конференции 9-11 ноября 2015 - Логистика и экономика ресурсоэнергобережения в промышленности. С. 63-68.
2. Аведьян Э.Д., Ле Т. Ч. Л. Нейронная сеть CMAC как альтернатива многослойной нейронной сети в задаче обнаружения DoS атак // Сборник научных трудов по материалам пятнадцатой национальной конференции по искусственному интеллекту с международным участием 3-7 октября 2016. Том 3. С 164-170.
3. Аведьян Э.Д., Ле Т. Ч. Л. Обнаружение DoS атак и их компонент на основе системы нейронных сетей CMAC// Сборник научных трудов по материалам XIV Всероссийская научная конференция «Нейрокомпьютеры и их применение» 15 марта 2016. С. 47
4. Ле Тхи Чанг Линь. Многослойная нейронная сеть в задаче обнаружения атак, представленных в современной базе данных UNSW-NB15// Сборник научных трудов по материалам III Инжиниринг и телекоммуникации – En&T 2016 29-30 ноября 2016. С. 163-164.
5. Ле Тхи Чанг Линь. Обнаружение атак с помощью многослойной нейронной сети по записям о сетевых соединениях современной базы данных UNSW-NB15// Сборник

научных трудов по материалам XV Всероссийская научная конференция «Нейрокомпьютеры и их применение» 14 марта 2017. С. 103.

6. Э.Д. Аведьян, Т.Ч. Ле Линь. Процедуры оптимального голосования в многоэкспертных бинарных системах // 60-я научная конференция МФТИ 20-26 ноября 2017. С. 140-141. https://abitu.net/public/admin/mipt-conference/FRKT.pdf
7. Т.Ч. Ле Линь. Оптимизация нейросетевой многоэкспертной системы обнаружения атак на современной безе данных UNSW-NB15// 60-я научная конференция МФТИ 20-26 ноября 2017. С.146-147. https://abitu.net/public/admin/mipt-conference/FRKT.pdf

8. Ле Тхи Чанг Линь, Дао Куанг Минь. Объединение метода случайного леса и многослойной нейронной сети для уменьшения числа признаков при обнаружении DoS-атак на основе базы данных UNSW-NB15// Сборник научных трудов по материалам IV международной конференции Инжиниринг и телекоммуникации – En&T 2017, 29-30 ноября 2017. С. 170-172.
9. Ле Тхи Чанг Линь. Сравнительный анализ оптимальных многоэкспертных бинарных систем в задаче обнаружения атак// Сборник научных трудов по материалам XVI Всероссийской научной конференции «Нейрокомпьютеры и их применение» 13 марта 2018. С. 194-195.

ЛИТЕРАТУРА

1. Lam L., Suen C.Y. Application of Majority Voting to Pattern Recognition: An Analysis of Its Behavior and Performance // IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans.–1997.–V. 27.–No. 5.–P.553-568.
2. Xu L., Amari Shun-ichi. Combining classifiers and learning mixture-of-experts // Encyclopedia of Artificial Intelligence.–2009.–P. 319-326.

Научное издание

Ле Тхи Чанг Линь

Оптимальные многоэкспертные бинарные системы голосования в модельных задачах обнаружения атак

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Федеральное государственное автономное образовательное учреждение
высшего образования
Московский физико-технический институт
(государственный университет)
141701,

Московская область, г. Долгопрудный
Институтский переулок, д. 9
https://mipt.ru