ОГЛАВЛЕНИЕ

Введение ... 3

1. Технологии искусственного интеллекта в задачах информационной безопасности ... 7

1.1. Подходы и направления развития искусственного интеллекта в ИБ ... 7

1.2. Методы и понятия машинного обучения в ИБ. Основные понятия ... 9

1.2.1. Классификация. Бинарная, многоклассовая, многозначная ... 11

1.2.2. Кластеризация ... 14

1.2.3. Регрессионный анализ ... 17

1.3. Автоматическая обработка текстов на естественном языке ... 19

Литература к разделу 1 ... 23

2. Алгоритмы машинного обучения в информационной безопасности ... 25

2.1. Особенности табличного представления исходных экспериментальных данных ... 25

2.1.1. Проблемы структуры табличных данных ... 25

2.1.2. Предобработка «исторических данных» ... 29

2.1.3. Предобработка несбалансированных данных ... 30

2.1.4. Использование разнородных типов данных ... 32

2.1.5. Способы отбора атрибутов исходного набора ЭД ... 33

2.1.6. Иные задачи машинного обучения ... 34

2.2. Однозначная классификация ... 35

2.2.1. Бинарная классификация ... 35

2.2.2. Многоклассовая классификация ... 37

2.3. Многозначная классификация ... 40

2.3.1. Особенности многозначной классификации ... 40

2.3.2. Методы решения задач многозначной классификации 41

2.2.3. Методы сведения многозначности меток классов к однозначному виду ... 43

2.3.4. Методы адаптации существующих алгоритмов ... 46

2.4. Метрики оценки эффективности классификации 48
2.4.1. Бинарная классификация 48
2.4.2. Многоклассовая класификация 52
2.4.3. Многозначная классификация 54
2.4.4. Некоторые известные функции потерь 57
2.5. Примеры алгоритмов классификации 61
2.5.1. Линейный классификатор 61
2.5.2. Логистическая регрессия 62
2.5.3. Наивный Байесовский классификатор 62
2.5.4. Алгоритм «K ближайших соседей» 63
2.6. Алгоритмы классификации на основе деревьев решений 65
2.6.1. Основные понятия 65
2.6.2. Classification and Regression trees 67
2.6.3. Алгоритм ID3 69
2.6.4. Алгоритм C4.5 70
2.6.5. Алгоритм CHAID 71
2.7. Ансамблевые алгоритмы 72
2.7.1. Методы композиции обучающихся алгоритмов 72
2.7.2. Бустинг 73
2.7.3. Бэггинг 74
2.7.4. Стекинг 76
2.8. Леса решений 77
2.8.1. Случайный лес. Принципы принятия решений о классификации за счет голосования 77
2.8.2. Isolation Forest 82
2.9. Метод усиления слабых моделей. Оценка устойчивости классификатора 88
2.9.1. Метод контрольных 90
2.9.2. Случайные подвыборки 90
2.9.3. Перекрестная проверка 91
2.10. Потоковая классификация 92
2.10.1. Основные понятия 92
2.10.2. Сценарии обработки потоковых данных 93
2.10.3. Дрейф концепта при потоковой классификации 94
2.10.4. Алгоритмы потоковой классификации 99
2.10.5. Эволюция алгоритмов потоковой классификации ... 105
2.10.6. Граница Хоэфдинга 106
2.10.7. Деревья Хоэфдинга 107
2.10.8. Особенности разработки программного обеспечения для классификации данных в потоковом режиме 107

2.10.9. Метрики оценки эффективности классификации потоковых алгоритмов ... 108
2.11. Классификация текстов ... 110
2.11.1. Постановка задачи классификации текста ... 110
2.11.2. Этапы и методы классификации текстов. Общие сведения ... 112
2.11.3. Индексация и предобработка наборов текстов (документов) ... 113
2.11.4. Извлечение термов. Взвешивание термов с использованием статистических мер. Взвешивание термов с использованием графа ... 116
2.11.5. Методы перевода текста в пространство признаков. Методы и технологии перевода слова в вектор фиксированной длинны ... 121
2.11.6. Метрики оценки качества и способы повышения точности классификации текста ... 124
2.11.7. Способы повышения точности классификации текста 125
Литература к разделу 2 ... 126
3. Алгоритмы классификации, базирующиеся на архитектуре искусственных нейронных сетей) ... 135
3.1. Базовые архитектуры ИНС ... 135
3.1.1. Многослойный персептрон ... 135
3.1.2. Сверточная нейронная сеть ... 136
3.1.3. Рекуррентная нейронная сеть ... 138
3.1.4. Архитектура LSTM ... 140
3.1.5. ИНС типа «автокодировщик» ... 142
3.1.6. Нейронные сети типа «трансформер» ... 144
3.2. Структура и режимы работы ИНС ... 144
3.2.1. Структура ИНС ... 144
3.2.2. Режим обучения ИНС ... 149
3.2.3. Рабочий режим ИНС ... 150
3.3. Классификация текстов на основе ИНС ... 151
3.3.1. Особенности построения полносвязных ИНС в задаче классификации текстов ... 151
3.3.2. Особенности построения ИНС типа РНС в задаче классификации текстов на примере РНС типа «Gated Recurrent Units» ... 154
3.4. Особенности построения современных ИНС ... 156
3.4.1. Комбинирование сверточных нейронных сетей и полносвязных слоев ... 156
3.4.2. Комбинирование ИНС типа «трансформер» и полносвязных слоев ... 158

3.4.3. Стратегии борьбы с переобучением в ИНС 159
3.4.4. Стратегии дообучения ИНС 160
Литература к разделу 3 161
4. Кластеризации в задачах ИБ 164
4.1. Место кластеризации в задачах ИБ 164
4.2. Принцип работы методов кластеризации 165
4.3. Неиерархические методы кластеризации 169
4.4. Итеративные методы кластеризации 169
4.4.1. Методы, базирующиеся на k средних (k-means) [33] .. 170
4.4.2. Метод CLOPE 172
4.4.3. Алгоритм PAM (partitioning around medoids) 173
4.5. Плотностные методы кластеризации 174
4.6. Модельные методы кластеризации 175
4.7. Сетевые методы кластеризации 176
4.8. Иерархические инкрементальные методы кластеризации 177
4.8.1. Алгоритм BIRCH 179
4.8.2. Метод ближайшего соседа 181
4.8.3. Минимальное покрывающее дерево 182
4.8.4. Метод COBWEB 182
4.8.5. Алгоритм объединения кластеров 183
4.9. Иерархические неинкрементальные алгоритмы 184
4.9.1. Метод Batch k-means 185
4.9.2. Online k-means 186
4.9.3. k-ближайших соседей 187
4.9.4. Алгоритм, основанный на определении степени обособленности векторов 187
4.9.5. Анализ отклонений 188
4.10. Самоорганизующаяся карта Кохонена 189
4.11. Сравнительный анализ методов кластеризации 192
4.11.1. Параметры сравнения 192
4.11.2. Результаты сравнения 193
4.12. Метрики оценки качества алгоритмов кластеризации 197
Литература к разделу 4 201
5. Программная реализация алгоритмов машинного обучения с использованием Python 204
5.1. Базы данных, используемых для решения задачи обнаружения компьютерных атак 205
5.1.1. База данных UNSW-NB15 205
5.1.2. База данных Kitsune 215

5.2. Настройка программной среды Python на примере PyCharm ... 219
5.2.1. Способ настройки среды через интерпретатор Anaconda ... 219
5.2.2. Способ настройки среды через python без сторонних интерпретаторов ... 222
5.2.3. Если проблема с установкой библиотек не устранена 225
5.2.4. Перечень библиотек, необходимых для решения задачи обнаружения атак ... 225
5.3. Предварительная обработка данных ... 226
5.3.1. Разделение выборки на тестовую и обучающую части 226
5.3.2. Выбор функции потерь ... 227
5.3.3. Масштабирование данных ... 227
5.3.4. Ввзаимное распределения атрибутов фрагмента исходных данных ... 228
5.4. Отбор количества и состава информативных признаков на примере базы данных UNSW-NB15 ... 231
5.4.1. Отбор информативных признаков с использованием моделей ... 233
5.4.2. Отбор признаков с использованием статистического подхода. Инструмент Feature Selector ... 236
5.4.3. Отбор признаков с использованием статистического подхода. Алгоритм SelectKBest ... 243
5.4.4. Использование метода главных компонентов ... 247
5.5. Сравнительный анализ методов отбора признаков и выбор состава и количества признаков на примере базы данных UNSW-NB15 ... 249
5.6. Бинарная классификация компьютерных атак на примере базы данных UNSW-NB15 ... 249
5.6.1. Перечнь алгоритмов классификации, используемых для классификации компьютерных атак ... 249
5.6.2. Оценка алгоритмов классификации ... 253
5.7. Многоклассовая классификация компьютерных атак на примере базы данных UNSW-NB15 ... 258
5.7.1. Особености задачи многоклассовой классификации .. 258
5.7.2. Алгоритмы на основе деревьев решений ... 260
5.7.3. Ансамблевые алгоритмы ... 262
5.7.4. Линейные классификаторы ... 269
5.7.5. Метрические классификаторы ... 275
5.7.6. Метод опорных векторов ... 277
5.7.7. Искусственные нейронные сети ... 278
5.7.8. Результаты многоклассовой классификации ... 279

5.8. Влияние многозначнных записей в базах данных на результаты многоклассовой классификации на примере базы данных UNSW-NB15 282
5.9. Обнаружение компьютерных атаки на примере базы данных Kitsune 296
5.9.1. Визуальная оценка данных 296
5.9.2. Проверка данных Kitsune на наличие выбросов, нечисловых и отсутствующих значений 297
5.9.3. Построение корреляционной диаграммы для сетевой атаки типа Botnet данных Kitsune 298
5.9.4. Алгоритмы бинарной классификации 300
5.9.5. Сравнительный анализ алгоритмов классификации . 308
5.10. Анализ и интерпретация результатов классификации на Python 310
Литература к разделу 5 313
Приложение 1. Исходный код примера «Kitsune» 317
Приложение 2. Лабораторный комплекс 324
Лабораторная работа № 1. Статистический анализ наборов данных 324
Лабораторная работа № 2. Исследование однослойных нейронных сетей н а примере моделирования булевых выражений 330
Лабораторная работа № 3. Оценка результатов классификации при помощи метрик TP, FP, FN, TN, Accuracy, Precision, AUC, F-мера, матрица ошибок 337
Лабораторная работа № 4. Алгоритмы машинного обучения с учителем для решения задач классификации 343
Лабораторная работа № 5. Алгоритмы машинного обучения с учителем для решения задач кластеризации 351
Лабораторная работа № 6. Применение однослойных нейронных сетей для решения задач регрессии на примере прогнозирования временного ряда методом скользящего окна 359
Лабораторная работа № 7. Применение библиотеки scikit-learn для решения задачи классификации на примере набора данных 363
Лабораторная работа № 8. Снижение размерности входных данных на примере алгоритмов PCA, SVD 366
Лабораторная работа № 9. Применение алгоритмов анализа категориальных данных. Моделирование топиков на основе алгоритма LDA 368
Литература к Приложению 2 375