ним для ускорения существующих алгоритмов решения родственных задач, связанных с нахождением максимальных полных подматриц 0,1-матрицы.

ЛИТЕРАТУРА

1. Биркгоф Г. Теория решеток. М.: Наука, 1984. 568 с.
2. Ganter B. and Wille R. Formal Concept Analyses: Mathematical Foundations. Springer Science and Business Media, 2012. 314 p.
3. Ganter B. and Obiedkov S. A. Conceptual Exploration. Berlin; Heidelberg: Springer, 2016. 315 p.
4. Kuznetsov S. O. and Obiedkov S. A. Comparing performance of algorithms for generating concept lattices // J. Exper. Theor. Artificial Intelligence. 2002. V. 14. No. 2. P. 189–216.
5. Mongush Ch. M. and Bykova V. V. On decomposition of a binary context without losing formal concepts // J. Siberian Federal University. Mathematics and Physics. 2019. No. 3. P. 323–330.
6. Быкова В. В., Монгуш Ч. М. Декомпозиционный подход к исследованию формальных контекстов // Прикладная дискретная математика. 2019. №. 44. С. 111–124.
7. Монгуш Ч. М., Быкова В. В. Программа FCASCorpus концептуального моделирования тувинских текстов методами анализа формальных понятий. Свид. о гос. регистрации программы для ЭВМ № 2018618907, выдано Федеральной службой по интеллектуальной собственности РФ, 2018.

УДК 519.7

DOI 10.17223/2226308X/12/63

ОБ ИСПОЛЬЗОВАНИИ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОВЕРКИ СТАТИСТИЧЕСКИХ СВОЙСТВ СИММЕТРИЧНЫХ КРИПТОГРАФИЧЕСКИХ АЛГОРИТМОВ

А. А. Перов

Рассматривается применение технологий машинного обучения к задачам криптографии, в частности проведению статистического анализа блочных шифров. Изложена идея адаптации шифртекстов к алгоритму модели нейронной сети Inception V3. Приведены результаты экспериментов.

Ключевые слова: криптография, машинное обучение, статистический анализ, раунд шифрования, итеративные блочные шифры.

Введение

Традиционно статистический анализ проводится с помощью тестов, определяющих степень случайности выходной последовательности [1] с применением методов математической статистики. Исследования показывают, что современные итеративные алгоритмы шифрования обеспечивают удовлетворительные статистические свойства на меньшем, чем полное, числе раундов. Использование технологий машинного обучения для решения подобных задач является новым направлением. Для разработки методики проведения статистического анализа была использована нейронная сеть Inception v3, обычно применяющаяся для распознавания и классификации графических образов. Inception v3 является свёрточной нейронной сетью, состоящей из 17 слоёв, обученной на большом количестве изображений из базы ImageNet.

1. Преобразование шифртекстов

Для решения задачи по обучению нейронной сети для статистического анализа выполнено преобразование зашифрованных сообщений в формат графических изображе-

ний. Для конвертации разработана утилита на языке C++, которая считывает в бинарном виде текстовый файл с зашифрованным сообщением и каждый байт зашифрованного текста записывает в качестве значения компонент палитры RGB, формируя изображения формата JPEG. При подаче на вход такой программе текстового файла с одним и тем же скопированным символом на выходе будет получено однотонное изображение. При подаче циклически повторяющегося набора символов изображение принимает характер узорчатого, что подтверждает корректность реализации.

Вторым подготовительным этапом является формирование базы шифртекстов различных алгоритмов на разном числе раундов шифрования. С помощью библиотеки «Униблок-2015» [2] получены шифртексты различных криптоалгоритмов на разном числе раундов. После завершения подготовительных этапов проведены эксперименты, ставшие основой методики проведения тестирования средствами технологии машинного обучения.

Эксперимент 1: были сгенерированы по 1000 шифртекстов алгоритма Simon на 3 и 30 раундах. Отличия на этих выборках видны человеку визуально: для последовательностей на 3 раундах очевидно прослеживаются зависимости, тогда как отображения 30-раундовых шифртекстов визуально имеют равномерное распределение. В данном эксперименте модель отличила 100 % поданных на тестовой выборке шифртекстов.

Эксперимент 2: сравнивались значения соседних раундов итеративного блочного шифра. В качестве обучающей выборки были поданы зашифрованные последовательности алгоритма Simon с 3 по 21 раунд (по 500 изображений на каждый). Модель должна была обнаружить, происходят ли изменения в структурах шифртекстов с каждым раундовым преобразованием. В отличие от эксперимента 1, разницу между соседними раундами определить визуально невозможно. На рис. 1 приведён график зависимости процента верных решений при различении раундов от их числа. Отмечается, что выборки шифра Simon на 3 и 4 раундах проходят одинаково малое число статистических тестов NIST, то есть для классического способа проведения статистического анализа выборки одинаковы, тогда как нейросеть отличила их более чем в 92 % случаев. Следует отметить, что после 15-го раунда модель правильно различала соседние раунды с вероятностью, стремящейся к 0,5, что может свидетельствовать о том, что качество выходной последовательности уже не улучшается и выборка обладает удовлетворительными статистическими свойствами.

Эксперимент 3: модель была обучена на выборках алгоритмов Speck и Present до 9-го раунда шифрования. Выбор этих шифров обусловлен схожим поведением при традиционном статистическом тестировании: на раундах до 9 оба алгоритма проходят примерно одинаковое число тестов. Несмотря на то, что алгоритмы по своей структуре выполняют разные преобразования над открытым текстом, модель способна успешно отличать друг от друга шифртексты до 7-го раунда (на первых шести раундах модель различила 100 % тестовых шифртекстов, к седьмому раунду процент понизился до 90,9 %). На 10-м раунде нейросеть при различении изображений приняла только 52 % правильных решений (статистические тесты определили именно эту границу минимального числа раундов, обеспечивающих удовлетворительные статистические свойства [2]). Найденное значение правильных решений модели может свидетельствовать о том, что оба алгоритма уже к 10-му раунду дают статистически качественный шифртекст.

Эксперимент 4: нейронная сеть была переобучена на изображениях шифртекстов разных шифров. На основании данных, полученных с помощью тестов NIST, на вход

Рис. 1. График зависимости процента правильных решений модели от числа раундов шифра Simon

сети было суммарно подано более 10000 шифртекстов всех алгоритмов библиотеки, зашифрованных на числе раундов, не обеспечивающих удовлетворительных статистических свойств. В качестве второй половины обучающей выборки использовалось также более 10000 шифртекстов, зашифрованных на полном числе раундов разными алгоритмами. В качестве тестовой выборки нейросеть получила более 2100 изображений, которые были случайным образом выбраны из базы сгенерированных шифртекстов. В 98 % случаев алгоритм машинного обучения верно определил степень случайности шифртекста.

Выводы

Широкое применение машинного обучения и, в частности, компьютерного зрения в самых разных областях жизни на сегодняшний день обусловлено высокими показателями результативности этих технологий. Описанные эксперименты подтвердили гипотезу о возможности применения нейронных сетей для решения задач анализа статистических свойств, так как полученные результаты согласуются с традиционными методами анализа посредством статистических тестов. Результаты экспериментов 2–4, в частности, свидетельствуют о том, что анализ статистических свойств для итеративных блочных шифров может проходить по разным сценариям, которые построены как на сравнении между собой шифртекстов одного алгоритма, зашифрованных на разном числе раундов, так и на сравнении зашифрованных последовательностей с эталонными, полученными в результате работы алгоритма на полном числе раундов.

ЛИТЕРАТУРА

1. Ryabko B. and Monarev V. Using information theory approach to randomness testing // J. Statistical Planning and Inference. 2005. V.133. No. 1. P. 95–110.

2. Перов А. А., Пестунов А. И. Статистическое тестирование современных итеративных блочных шифров с помощью программной библиотеки «УНИБЛОКС-2015» // Инновации в жизнь. 2016. № 2. С. 89–97.

УДК 519.6

DOI 10.17223/2226308X/12/64

СПОСОБ РЕШЕНИЯ НЕДООПРЕДЕЛЁННЫХ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ НАД $GF(2)$ С ИСКАЖЁННЫМИ ПРАВЫМИ ЧАСТЯМИ И ОГРАНИЧЕНИЕМ НА МАЛЫЙ ВЕС РЕШЕНИЯ

Н. Ю. Руменко, А. В. Костюк

Рассматриваются недоопределённые случайные системы линейных булевых уравнений с искажёнными правыми частями, истинное решение которых имеет малый вес Хемминга. Экспериментально показывается, что для малых вероятностей искажения такие системы могут быть эффективно решены применением алгоритмов декодирования по информационным множествам.

Ключевые слова: случайные системы линейных булевых уравнений, декодирование по информационным множествам.

Рассмотрим систему из фиксированного числа $m < n$ линейных булевых уравнений (СЛУ)

$$Ax = b = Ax_0 \oplus \xi, \quad (1)$$

где A — случайная двоичная матрица размера $m \times n$; $\xi^T = (\xi_0, \dots, \xi_{m-1})$ — случайный двоичный вектор ошибок, координаты которого независимы и принимают значения с вероятностями $P\{\xi_i = 1\} = 1 - P\{\xi_i = 0\} = p$, $i \in \{0, \dots, m-1\}$, $p \in [0, 1/2]$; x_0 — вектор истинного решения, $\|x_0\| = w$, $\|\cdot\|$ — вес Хемминга.

Для малых значений w известны алгоритмы решения таких систем, основанные на переборе возможных значений истинного решения [1, 2].

В алгоритме максимума правдоподобия [1] непосредственно вычисляются все возможные векторы y длины n веса w и в качестве решения выдаётся вектор с наименьшим значением $\|Ay\|$. Алгоритм имеет асимптотическую сложность $O((1-2p)^{-2}n^w \log n)$ при использовании $O(1)$ бит дополнительной (помимо хранения системы уравнений) памяти [2].

В работе [2] рассматривается вычислительно более эффективный алгоритм на основе метода «встречи посередине» с емкостной сложностью $O(n^{w/2})$ и вычислительной сложностью $O(n^{w/2}l((m-t) \log n^{w/2} + wm))$, где t — пороговое значение $\|Ay\|$; l — число итераций.

Вычислительная сложность алгоритмов данного класса более всего зависит от параметра w , что делает их малоприменимыми для слабоискажённых систем с достаточно большим весом истинного решения.

С другой стороны, матрицу A можно рассматривать как проверочную матрицу некоторого случайного линейного блочного кода $M(n, n-m)$, для которого вектор b является синдромом ошибки веса w , причём каждый бит синдрома дополнительно искажён с вероятностью p . Задача решения системы (1), таким образом, состоит в декодировании случайного кода по искажённому синдрому.

Составим расширенную систему

$$Ch = b, \quad (2)$$