УДК 004.056:004.89

АЛГОРИТМ ПРЕДОБРАБОТКИ ДАННЫХ ДЛЯ НЕЙРОСЕТЕВОЙ СИСТЕМЫ ОПРЕДЕЛЕНИЯ АВТОМАТИЧЕСКОГО ПОДБОРА ПАРОЛЯ

Мансур Али Махмуд
Донской государственный технический университет, Ростов-на-Дону, Российская Федерация
asklabious@gmail.com

Рассматривается одна из важных проблем информационной безопасности, которая заключается в использовании программ автоматического подбора пароля. Предлагается подход к решению данной проблемы с применением инструментов интеллектуального анализа данных, в частности, на основе искусственных нейронных сетей. В качестве основы для реализации предлагаемых средств разработан алгоритм предобработки данных для нейросетевой системы определения автоматического подбора пароля как наиболее важный шаг, предваряющий процесс анализа данных. Подробно описывается способ расчёта статистических параметров, необходимых для формирования входных векторов нейронной сети, используемой для последующего обнаружения шаблонов и закономерностей, позволяющих определить использование автоматических средств перебора паролей.

Ключевые слова: интеллектуальный анализ данных, нейронные сети, грубая сила.

UDC004.056:004.89

DATA PRE-PROCESSING ALGORITHM FOR THE NEURAL NETWORK SYSTEM FOR DETERMINING AUTOMATIC PASSWORD SELECTION

Mansour Ali Mahmoud
Don State Technical University, Rostov-on-Don, Russian Federation
asklabious@gmail.com

The article is devoted to one of the important problems of information security, which is the use of automatic password selection programs. It proposes an approach to solving this problem using data mining tools, in particular, the ones which are based on artificial neural networks. As a basis for the implementation of the proposed tools, a data pre-processing algorithm has been developed for the neural network system for determining automatic password selection as the most important step anticipating the data analysis process. The paper describes in detail the method of calculating the statistical parameters necessary for the formation of the input vectors of the neural network used for the subsequent detection of patterns that determine the use of automatic brute-force tools.

Keywords: data mining, neural nets, brute-force.

Введение. Применение программных средств автоматического подбора пароля представляет серьезную угрозу информационной безопасности, возникающую при реализации атаки методом «грубой силы» (bruteforce) [1, 2]. Такие атаки имеют характерные параметры и свойства, которые позволяют обнаруживать их с использованием средств интеллектуального анализа данных, в частности, на основе искусственных нейронных сетей [3, 4].

Важным этапом, предшествующим непосредственно обучению нейронной сети, является предобработка исходных данных, позволяющая представить их в форме, пригодной для анализа, с целью формирования оптимальной структуры нейронной сети [5]. При этом, с одной стороны, стремятся сократить размерность входных данных, приведя её к одинаковым значениям для разных элементов выборки, а с другой — сохранить и по возможности акцентировать присутствующие в данных закономерности [6]. Данный этап включает в себя подготовку источника данных и

настройку векторов нейронной сети, которые будут использоваться для обучения и тестирования создаваемой системы.

В качестве критериев автоматического подбора пароля могут быть использованы некоторые статистические характеристики процессов авторизации пользователей какой-либо системы [7]. К ним относится в первую очередь среднее время между попытками входа, а также параметры отклонения от него временных промежутков между попытками входа. Признаком для определения автоматического подбора должно быть большое количество неудачных попыток входа и равные промежутки между ними. Цель данной работы — анализ методов, позволяющих эффективно использовать процесс предварительной обработки данных (с применением различных средств разработки) и дающих возможности для применения искусственных нейронных сетей для решения задач в области информационной безопасности.

Требования к алгоритму. Исходными данными для создания алгоритма, рассматриваемого в статье, являются журналы подсистемы аудита безопасности, использующиеся во многих информационных системах. Они, как правило, включают в себя записи о наиболее важных с точки зрения безопасности действиях пользователей, времени их совершения и результате (рис. 1). Результатом работы алгоритма должна быть таблица, столбцы которой представляют собой входные векторные элементы нейронной сети исследуемой системы [8, 9]. Каждая такая строка должна включать:

— количество неудачных попыток входа в течение минуты,

— среднее время между неудачными попытками входа в систему,

— стандартное отклонение интервалов от среднего времени между попытками авторизации.

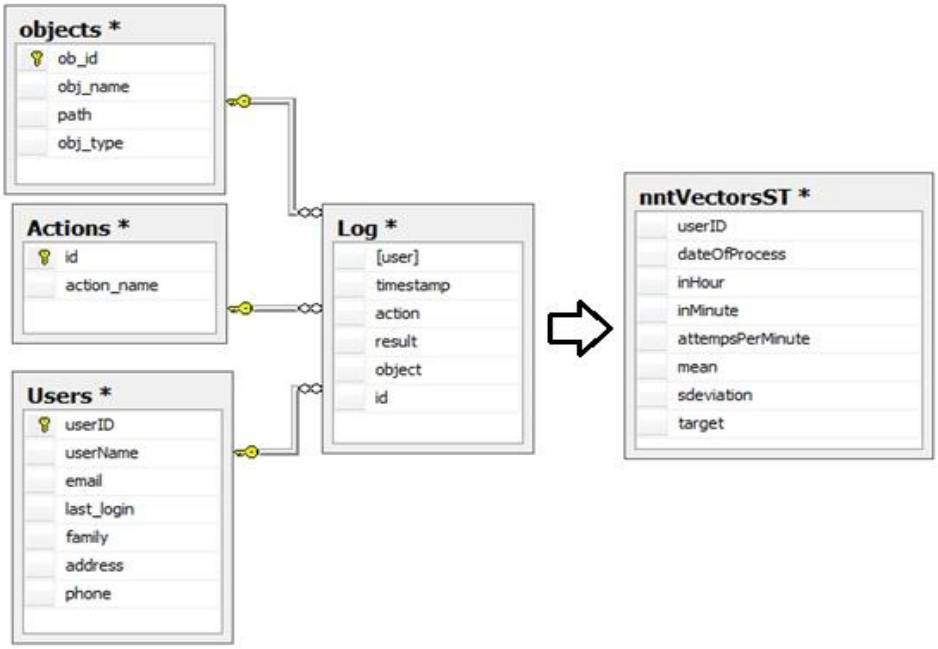

Рис. 1. Схемы таблиц исходных данных и результата

Последовательность действий алгоритма. Предлагаемый алгоритм предобработки данных включает в себя несколько этапов, подробное описание которых приведено ниже.

1. Выбор данных. На этом этапе для экономии вычислительных ресурсов и памяти из таблицы исходных данных исключаются столбцы и строки, которые не будут использоваться в дальнейших расчётах. Данный этап включает в себя две задачи: сначала идентифицировать записи, содержащие сведения о попытках авторизации на основе значения кода операции, а после — вы-

брать столбцы, которые нужны для вычисления необходимых статистических значений, а именно столбец результата (успешный/неудачный), идентификатор пользователя и время выполнения действия.

2. Агрегация данных. На этом этапе производится группировка данных, относящихся к одному пользователю, но находящихся в разных строках таблицы. Этот процесс требует разделения метки времени (timestamps) на элементы, которые её составляют — дата и время (часы, минуты, секунды) — для того, чтобы пользовательские данные могли быть сгруппированы по любому из этих элементов времени. Сама группировка данных выполняется одной из агрегатных функций языка SQL (avg, sum и т.д.) и оператора groupby.

3. Расчет статистических значений. На данном этапе производится вычисление интервала (разницы во времени) между каждыми двумя последовательными попытками входа за одну минуту. На практике интервалы вычисляются с помощью SQL-запросов, а для упрощения понимания механизма их расчёта ниже приведён его алгоритм.

Шаг 1. Подсчет количества пользователей в системе:
КП=количество пользователей,
установить *С* в качестве счетчика для пользователей.

Шаг 2.Пока есть пользователи, для которых *С*<*КП*:
для пользователя *С*: найти общее количество попыток входа,
КЗ=количество записей пользователя,
установить *j* в качестве счетчика для записей.

Шаг 3.Пока есть записи для пользователя *С*, у которого *j*<*КЗ*:
за конкретный период времени «минута *i*»:
найти количество попыток входа в течение конкретной минуты (*i*),
N=количество попыток входа в течение минуты (*i*),
установить *i* в качестве счетчика для минут,
если *N*=1, то разность равна нулю (интервал=0),
если существует более одной записи *N*>1 и пока *i*<*N*:
рассчитать разницу между каждой последовательной попыткой входа,
Интервал=Следующая Попытка – Предыдущая Попытка,
перейти к следующей минуте (*i*+1).

Шаг 4. Масштабирование. Целью этого процесса является получение результирующих значений в интервале [0, 1] для соответствия входам нейронной сети. Для этого каждый результат (интервал) делится на 100 000. Причина, по которой выбрано 100 000, заключается в том, что разница рассчитывается в миллисекундах, поэтому самый длинный номер выхода будет состоять из пяти цифр.

Шаг 5. Сохранить этот интервал в таблице входных векторов нейронной сети.

Схема этого алгоритма показана на рис. 2.

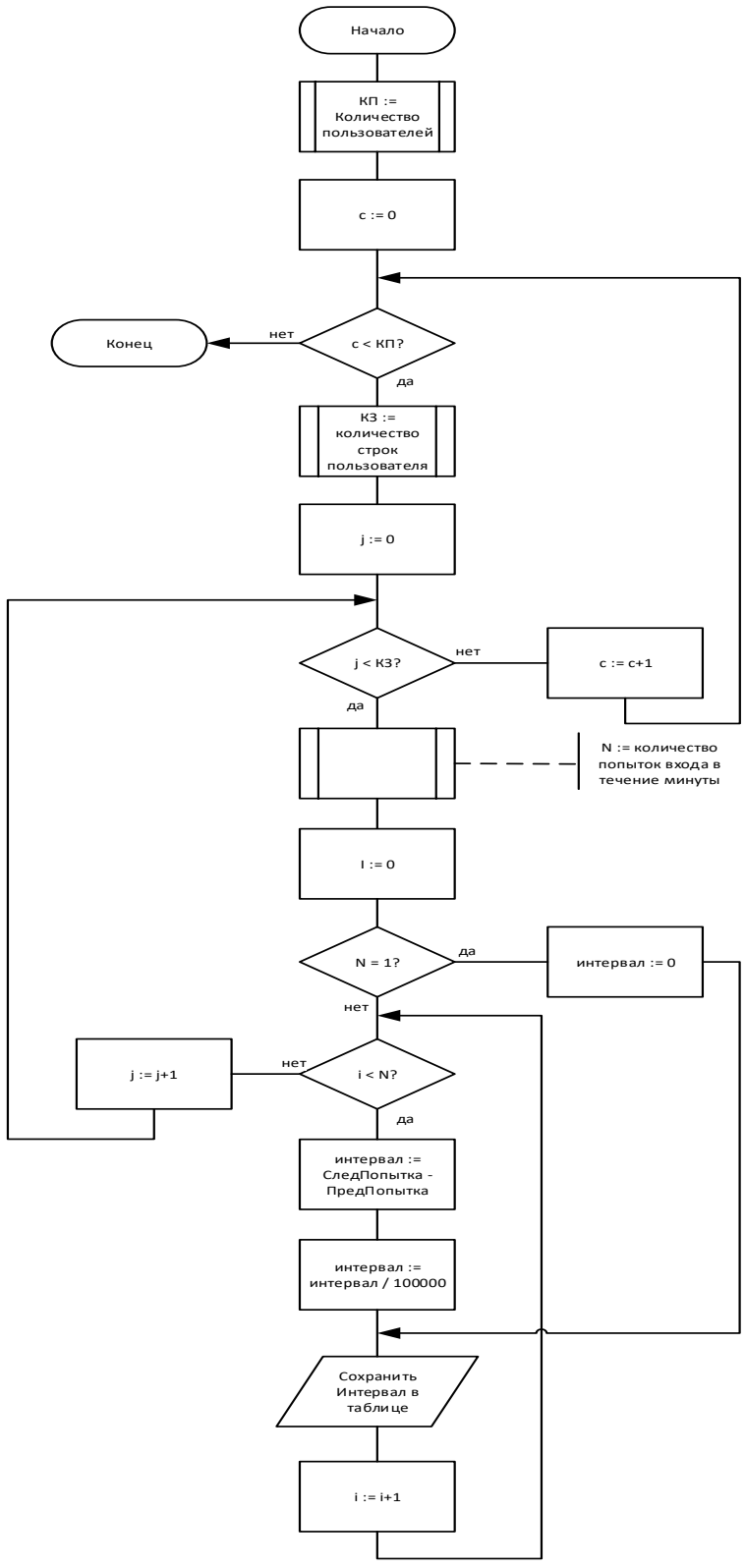

Рис. 2. Алгоритм расчёта временных интервалов

Вычисление статистических показателей. После вычисления интервалов между каждой последовательной попыткой входа легко вычислить среднее арифметическое и стандартное отклонение для этих периодов в течение указанного времени. Среднее значение вычисляется по формуле:

$$\overline{\mathrm{T}} = \frac{1}{N}\sum_{i=1}^{n} \mathrm{T_i}\ ,$$

где $\bar{T}$ — среднее значение интервалов в течение минуты,

N —количество попыток входа пользователя в течение минуты,

T_i — интервал между каждыми двумя последовательными попытками входа в течение минуты, где i обозначает номер пользователя.

Стандартное отклонение можно выразить формулой:

$$S_i = \sqrt{\frac{\sum_{i=1}^{N}(T_i - \bar{T})^2}{N_i}},$$

где T_i — интервал между каждыми двумя последовательными попытками входа в течение минуты, где i обозначает номер пользователя,

$\bar{T}$ — представляет среднее значение этих интервалов в течение минуты,

N — количество попыток входа пользователя в течение минуты.

Полученные таким образом значения полностью соответствуют структуре таблицы результатов предобработки, показанной на рис. 1, и готовы для анализа нейронной сетью.

Заключение. Рассмотренные в статье методы позволяют эффективно реализовывать процесс предварительной обработки данных с использованием различных средств разработки, в том числе языка SQL, и открывают возможности для дальнейшего применения искусственных нейронных сетей как одного из средств интеллектуального анализа данных для решения задач в области информационной безопасности.

Библиографический список

1. TruptibenD. Brute-force Attack “Seeking but Distressing”. / D.Truptiben // International Journal of Innovations in Engineering and Technology. — 2013. — pp. 75-78.

2. Knudsen L. Brute force attacks / L.Knudsen, M. Robshaw //The Block Cipher Companion. — Berlin: Springer, 2011. — pp. 95-108.

3. Агеев, С. А. Методы интеллектуального анализа данных для управления рисками информационной безопасности в защищенных мультисервисных сетях специального назначения / С. А. Агеев // Автоматизация процессов управления. — 2015. — № 2. — С. 42–49.

4. Али, М. Особенности применения средств объектно-ориентированного программирования для реализации многослойных искусственных нейронных сетей прямого распространения / М. Али // Сборник статей Международной научно-практической конференции «Автоматизация: проблемы, идеи, решения». — Уфа : Омега сайнс, 2017. — С. 107–109.

5. Аникин, И. В. Технология интеллектуального анализа данных для выявления внутренних нарушителей в компьютерных системах / И. В. Аникин // Научно-технические ведомости СПбГПУ. — 2010. — № 6 (113). — С. 112–117.

6. Анализ данных и процессов : Учеб. пособие. 3-е изд., перераб. и доп. / А. А Барсегян [и др.]. — Санкт-Петербург : БХВ-Петербург, 2009. — 512 с.

7. Методы и модели анализа данных: OLAP и Data Mining / А. А. Барсегян [и др.]. — Санкт-Петербург : БХВ-Петербург, 2004. — 336 с.

8. Новиков, Ф. А. Дискретная математика для программистов / Ф. А. Новиков. — Санкт-Петербург : Питер, 2000. — 364 с.

9. Роднин, А.В. Концепция применения интеллектуального анализа данных в средствах защиты информации баз данных / А. В. Роднин, В. Ю. Турчик // Физика. Технологии. Инновации: сборник научных трудов. — 2015. — Вып. 1. — С. 263–269.