БЕЗОПАСНОСТЬ И ШИФРОВАНИЕ ДАННЫХ ПРИ ПОМОЩИ НЕЙРОННЫХ СЕТЕЙ

Nichita ȚURCAN

Universitatea Tehnică a Moldovei

Abstract*:Конфиденциальность информации – одна из важнейших задач на сегодняшний день в информационном поле. Несмотря на то, что нейронные сети изначально не предназначены для хранения информации – применение НС в данном сегменте обладает большим потенциалом.*

Keywords*: internet, neural network, data storage, applied informatics, new technologies.*

Введение

В последнее время, в научном сообществе наблюдается небывалый всплеск интереса к нейронным сетям. Как уже было доказано на практике – спектр распространения и применения нейронных сетей – достаточно широк: начиная от бизнеса и медицины, заканчивая машиностроением и геологией. Нейронные сети вошли в практику везде, где нужно решать задачи прогнозирования, классификации или управления. Повышенному интересу нейронные сети обязаны по нескольким причинам:

Неограниченные возможности. Нейронные сети - мощный инструмент моделирования, позволяющий создавать и решать чрезвычайно сложные зависимости. Стоит обратить внимание, что нейронные сети – не линейны, а на протяжении многих лет линейное моделирование было основным методом. Существует множество задач, где решение при помощи линейной аппроксимации не дает положительных результатов - из этого вытекает, что линейные модели в таких задачах работают крайне неудовлетворительно. Еще одним плюсом стал тот факт, что нейронные сети справляются с моделирование линейных зависимостей с большим числом переменных.

Легкость в применении. Один из основных плюсов в нейронных сетях – это самообучаемость. Достаточно подобрать входные данные и запустить алгоритм обучения, который используя примеры будет автоматически воспринимать любую структуру данных. Пользователь при этом, должен, обладать минимальными навыками о том, как следует отбирать и подготавливать данные, выбирать нужную архитектуру сети и интерпретировать результаты, однако уровень знаний, необходимый для успешного применения нейронных сетей, гораздо скромнее, чем, например, при написании алгоритма на любом из высокоуровневых языков программирования.

Нейронные сети привлекательны тем, что могут быть использованы для решения, казалось бы, стандартных задач, но с абсолютно другим подходом. На сегодняшний день, одним из перспективных направлений является изучение ассоциативной памяти на нейронных сетях. Которое, в свою очередь, является мощным оружием в арсенале специалиста по прикладной информатике.

Постановка задачи и ее решение

Как и упоминалось ранее, возможности нейронных сетей практически безграничны. Существует множество инструментов для классификации каких-либо объектов по входным данным в нейронных сетях. Но стоит изменить алгоритм взаимодействия с сетью и можно хранить в ней длинные строковые или бинарные данные, получаемые по ключ-паролю, что в свою очередь позволит повысить уровень криптоустойчивости данных.

В тоже время, в виде матрицы весов нейронной сети возможно хранить конфиденциальные данные, начиная от обычных паролей и пин-кодов для банковских кард, заканчивая визуальными и голосовыми паролями.

Обзор исследований в области нейронных сетей

В литературе можно встретить множество публикаций, посвященных научным исследованиям возможности использования НС при решении задач информационной безопасности. К сожалению, подавляющее большинство таких исследований носит теоретический характер и не доведены до широкого практического внедрения. Ниже приведем примеры.

1. Обнаружение вторжений. Исследованию указанной проблемы посвящено самое большое число публикаций. В последнее время стали популярны методики построения обучающей модели для нейросети на базе СОВ с открытым кодом «Snort».

В таких работах, как правило, в качестве архитектуры НС рассмотрены различные виды сетей адаптивного резонанса (классификаторы Карпентера-Гроссберга) и обобщенно-регрессионная сеть.

Моделирование испытаний по требованиям безопасности. В работах с моделированием сетевых атак, приведено авторское исследование Тумояна Е.П по оценке и планированию испытаний программных средств защиты. Исследование показала эффективность применения технологий нейронных сетей при испытаниях интегрированных программных изделий, включающих компоненты с открытым кодом, когда рост степени технологической безопасности имеет немонотонный характер (см. рис.1). В частности, в работе показано, что лучше всего с поставленной задачей справились простая 4-слойная нейронная сеть, а также нейронные сети с обходными соединениями.

Использование технологии нейронных сетей в биометрии. Одно из популярных направлений развития нейросетевых технологий связано с биометрией. Причем, речь не только об аутентификации отпечатков пальцев или сетчатки глаза, но и о поведенческих характеристиках, например, почерк.

2.Нейронные сети в криптографии и стеганографии. Хоть использование нейросетевых технологий в криптографии не сильно обсуждаемая тема, но есть несколько интересных публикаций в данной тематике.

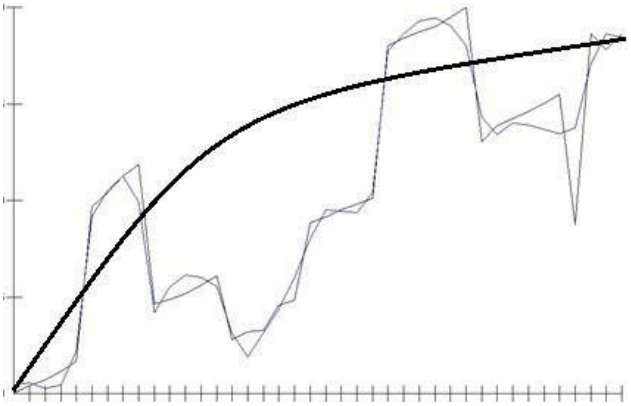

Рис. 1. Модели оценки уровня технологической безопасности программ (НС-модель показана тонкой линией, экспоненциальная – толстой)

В основном применение НС связано с криптосистемами, когда в качестве открытого ключа криптосистемы используется ключ Меркла-Хэллмана.

3. Другие применения аппарата нейронных сетей в области информационной безопасности. Например, в [6] предложен нейросетевой подход к иерархическому представлению коммуникационной сети, что может быть актуальным при исследовании безопасности, например, социальных сетей. Другим примером является использование нейронной сети для тестирования CAPTCHA [7].

Кроме указанных примеров теоретических изысканий, в открытых источниках можно встретить упоминание практических реализаций, например:

- системы распознавания, которые разрабатывает лаборатория GoogleX;
- работы ряда фирм, ориентированных на борьбу с мошенничеством в банковской сфере.

Что касается первого примера, то следует отметить технологический стартап Vicarious, в рамках которого, как заявлено, решена задача прохождения CAPTCHA-тесты, в том числе наиболее популярной в современном интернете - reCAPTCHA. Исследователям удалось достичь 90% точности распознавания CAPTCHA от Google, Yahoo, PayPal, Captcha.com и других проектов. Это исследование показывает, что современные CAPTCHA уже не эффективны в качестве теста Тьюринга. В рамках другого проекта ученые лаборатории Google X, работающие над созданием экспериментальной компьютерной нейронной сети, научили ее самостоятельно распознавать в видеопотоке морфологические объекты, например, кошек.

Подбор метода

Преступая к разработке нейросетевого решения, как правило сталкиваются с выбором оптимальной архитектуры нейронной сети. В зависимости от использования того или иного типа нейронных сетей – результаты могут разниться.

Задачи хранения конфиденциальной информации можно решать с помощью нейронных сетей следующих типов: Нейронные сети Хопфилда, Рециркуляционные нейронные сети и Ассоциативные нейронные сети.

Нейронную сеть Хопфилда можно использовать как ассоциативную память, ровно, как и сеть Кохонена, при необходимости, вышеуказанные архитектуры можно использовать как некое запоминающее устройство (по кластеру определяются наиболее вероятные значения исходных сигналов).

Нейронные сети Хопфилда при восстановлении образа стремится к своему равновесному состоянию, которое и будет сохраненным ею образом. Однако для запоминания паролей такую сеть использовать не получится, в силу особенностей ее архитектуры.
Автоассоциативные нейронные сети, ровно, как и метод Коско не подходят для хранения информации, т.к. восстанавливают на выходе то, что было подано на входе. При помощи автоассоциативных нейронных сетей реализуют нелинейный метод, либо используют для минимизации входного вектора.

Но точно такого же результата можно добиться, используя широкий класс рекуррентных нейронных сетей, классическим примером которых служит сеть Элмана, при этом проблема устойчивости отпадает, а на весовые коэффициенты не накладываются такие жесткие условия, благодаря чему сеть обладает большей емкостью. Кроме того, рекуррентные нейронные сети могут описывать конечный автомат, при этом не теряя всех преимуществ искусственных нейронных сетей.
В конечном итоге, за основу был взят однослойный персептрон. Для которого необходимо было создать прототип класса расчёта по матрице весов. Особенность персептрона - это 256 (по количеству символов в ASCII) нейронов во входном и скрытом слоях. На выходе же - один нейрон, который выдает номер нейрона, на котором получилась наибольшая сумма.

Значения весов сохраняется в матрице alpha [i, j], где i – номер нейрона скрытого уровня, а j – номер связи. На вход подается строка из 5 символов (наиболее оптимальный вариант). Таким образом, вычисление выходного символа сводится к двум функциям: преобразование входной строки в массив входных нейронов и выделение выходного символа исходя из входных нейронов и матрицы весов.

Предварительные преобразования

Если взять за основу тот факт, что чем правее находится символ в исходной строке, тем более он влияет на последующие символы, то на выходе получается простая функция преобразования строки в массив «сигналов» входного слоя. Рассмотрим на простом примере: для входной строки «17345» существует некий массив k, в котором k['1'] = 1; k['7'] = 2 и т.д.
Из этого, расчет выхода сводится к циклу вида:

```
public int[] AlphaProcessing(int[] _in)
{
    int[] o = new int[256];
    for (int i = 0; i <= 255; i++)
        for (int j = 0; j <= 255; j++)
            o[i] = alpha[i, j] * _in[j] + o[i];
    return o;
}
```

Где входящий массив _in[] есть выходной массив k[] из предыдущей функции.

Вслед за расчетом сигналов нейронов скрытого уровня (выходной массив o[]), отбирается максимальный, он и будет искомым выходным символом.

Кодирование информации

Цель - обучить нейронную сеть предсказывать данные по ключ-паролю. В качестве прогнозирующего метода был использован метод окон (применяемый при прогнозирование валютных колебаний). Алгоритм обучения нейронной сети будет выглядеть следующим образом:
1) Обработка и вычисление 5 первых символов строки.
2) Сравнение со следующим символом входящей строки и при необходимости проводится обучение/перестановка весов.
3) Пункты 1 и 2 повторяются со смещением на один символ в право до конца входной строки.

Таким образом, необходимо используя правильную последовательность, составить строку для начального обучения.
paswd+++Hello world! It's me!\0
Структура:
1) paswd – ключ-фраза (количество символов неограниченно)
2) +++ – три проверочных символа

3) Hello world! It's me! – кодируемая фраза
4) \0 – конец кодируемой фразы

Декодирование информации

Декодирование должно производится с единственной входной информацией в виде ключ-пароля.

Для решения поставленной задачи, необходимо проверить данные на корректность. Для этого требуется восстановление проверочных символов, указанных в структуре. Алгоритм работы сводится к следующим шагам: Подача на вход 5 символов пароля. Получение первого символа. Формирование нового входа - четыре символа пароля, а также символ проверки. Получение второго символа и сравнение с первым. Аналогичная процедура повторяется и с третьим символом. В случае успеха, получившееся для последующих итераций входная строка будет состоять из 2 символов пароля и трех символов проверки.

Следующий шаг – это восстановление информации. Необходимо проводить итерации до тех пор, пока алгоритм не дойдет до стоп- символа. В каждой последовательной итерации во входной строке отсекается первый символ и дополняется предыдущим выходным символом. Каждый выходной символ дополняется к строке вывода. Таким образом, дойдя до конечного символа, получаем закодированную строку на выходе.

Хранение информации в нейронной сети

Исходя из проведённого анализа, было принято решение хранить полученную информацию в виде бинарного файла с записанным в него сериализованным массивом весов (alpha[,]). Причина прежде всего кроется в использовании C# для написания алгоритма.

На первый взгляд может показаться избыточным хранить небольшие строки в объемном файле. Однако, тестовую строку, используемую в качестве примера, нейронная сеть совершила 23 итерации при обучении. Для хранения больших объемов информации, более 100 символов понадобилось порядка 2500 итераций. В итоге, данный алгоритм демонстрирует довольно высокую криптоустойчивость, при условии, что нейронную сеть необходимо постоянно обучать и испытывать.

Библиография:

1. Ф. Вассерман. Нейрокомпьютерная техника: Теория и практика. — М.: «Мир», 1992.
2. Саймон Хайкин. *Нейронные сети: полный курс = Neural Networks: A Comprehensive Foundation*. — 2-е изд. — М.: «Вильямс», 2006. — С. 1104. — ISBN 0-13-273350-1.
3. Т. Кохонен. *Ассоциативные запоминающие устройства = Content-Addressable Memories.* — М.: "Мир", 1982.
4. Головко В.А. *Нейронные сети: обучение, организация и применение*. М.: Издательство "Радиотехника", 2001.
5. https://ru.wikipedia.org/wiki/Ассоциативная память на нейронных сетях.
6. Басараб М.А., Вельц С.В. Нейросетевой подход к иерархическому представлению компьютерной сети в задачах информационной безопасности// Инженерный журнал: наука и инновации. 2013. № 2 (14). Режим доступа: http://engjournal.ru/catalog/it/security/534.html (дата обращения 12.11.2015).
7. Vicarious Solves CAPTCHA. Режим доступа: http://news.vicarious.com (дата обращения: 28.01.2014