

А.М. ДАВИДЕНКО, канд. техн. наук,
С.Я. ГИЛЬГУРТ, канд. техн. наук,
О.О. ВИСОЦЬКА,
А.А. КОЧУРКОВ,
Ю.О. ЧЕРНОВА

ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ПРОГРАМИ ДЛЯ АВТЕНТИФІКАЦІЇ КОРИСТУВАЧІВ КОМП'ЮТЕРНОЇ СИСТЕМИ ЗА КЛАВІАТУРНИМ ПОЧЕРКОМ ЗА ДОПОМОГОЮ ІМОВІРНІСНОЇ НЕЙРОННОЇ МЕРЕЖІ

Розроблена та експериментально досліджена програма для визначення ефективності застосування імовірнісних нейронних мереж для рішення задачі автентифікації користувачів комп'ютерної системи за клавіатурним почерком. Досліджувався вплив на її ефективність різних параметрів, наприклад відсоток помилок при наборі тексту, кількість грубих помилок у значеннях ознак в навчальних зразках. Запропоновано засоби усунення труднощів, які пов'язані по-перше з часовими витратами, а по-друге з підвищеними вимогами до обчислювальної техніки.

Worked out and experimentally studied program for measuring the effectiveness of application of probable neuron networks for solving problems of authentication of computer system users according to keyboard handwriting. The influence of various parameters on its effectiveness have been studied, for example, the percentage of errors while typing the text, the number of rough mistakes in the meanings of characteristics in learning samples.

Постановка проблеми та мета досліджень. Для визначення впливу найбільш характерних параметрів, які впливають на ефективність (E_f) застосування ймовірнісних нейронних мереж [1, 2] для рішення задачі автентифікації користувачів комп'ютерної системи, або іншими словами, для визначення ймовірності правильного розпізнавання користувачів системи було спочатку накопичений необхідний обсяг інформації, після чого було проведено ряд експериментів для визначення ефективності застосування ймовірнісних нейронних мереж для рішення задачі автентифікації користувачів комп'ютерної системи. Для цього використовувалася спеціально створена програма [3], яка написана на мові Builder Borland C++. Всі експерименти можна розділити на три етапи:

1. На першому етапі [4] аналізувалися дані для всіх користувачів, при цьому розглядалися всі навчальні зразки, тобто час набору символів кожного слова. На цьому етапі визначався вплив наступних параметрів на ефективність застосування імовірнісної нейронної мережі для рішення поставленої задачі:

◆ Найвність усереднення навчальних даних (usr), при якому усереднювалися ознаки із кожних десяти записів. При цьому збільшувалася точність даних, але зменшувалася їх кількість.

◆ Наявність відбору за словами (*ots*), тобто "невідомі" екземпляри порівнювалися з усіма навчальними даними або тільки з такими ж словами. Якщо відбір не проводився, тоді неточною була тільки перша ознака, а якщо проводився відбір, цих неточностей не було, але кількість навчальних даних зменшувалося в 10 разів, тому що в наборі було 10 слів.

- ◆ Кількість ознак (n).
- ◆ Кількість навчальних даних для кожного користувача (z_t).
- ◆ Ширина функції активності (σ).

2. На другому етапі [4] досліджувався вплив на ефективність застосування імовірнісної нейронної мережі для рішення поставленої задачі таких параметрів, як відсоток помилок при наборі тексту (Och) (уважність користувачів при роботі за клавіатурою), амплітуда (A_{Och}) відсотку помилок у користувачів між собою. Крім того, визначався вплив цих факторів на залежність ефективності застосування імовірнісної нейронної мережі для рішення задачі автентифікації користувача комп'ютерної системи від розглянутих раніше параметрів.

3. На третьому етапі [5] досліджувався вплив на ефективність застосування імовірнісної нейронної мережі для рішення поставленої задачі такого параметра, як кількість грубих помилок у значеннях ознак в навчальних зразках. Також досліджувалося, який з розглянутих алгоритмів виключення навчальних зразків із грубими помилками кращий. Після вибору кращого алгоритму, визначався максимально допустимий відсоток відхилення ($|u_{ijt} - Sr_{it}|$) кожної ознаки від її середнього значення (Sr). Крім того, визначався вплив цього фактора на залежність ефективності використання імовірнісної нейронної мережі для рішення задачі автентифікації користувача комп'ютерної системи від розглянутих раніше параметрів.

При проведенні експериментів виникли труднощі, які пов'язані по-перше з часовими витратами, а по-друге з підвищеними вимогами до обчислювальної техніки, яка використовується. Ці труднощі пояснюються, тим, що імовірнісні нейронні мережі дуже вибагливі у відношенні ресурсів.

Для накопичення необхідної кількості навчальних даних потрібні великі часові витрати. Так, якщо вважати, що мінімальна необхідна кількість навчальних даних (z_t) – 1000 та по 1000 екземплярів використовувати в якості невідомиз зразків, тоді для накопичення даних для одного користувача системи необхідно приблизно 4 – 5 годин, а якщо тестувати мережу за умови що в системі 20 користувачів, тоді необхідно 80 – 100 годин. Для зменшення часових витрат адміністратора, який повинен контролювати процес накопичення навчальних даних, в програмі передбачена можливість накопичення навчальних даних на різних комп'ютерах, але при цьому не знижуються часові витрати користувачів системи та витрати, які пов'язані з використанням комп'ютерів.

При аналізі накопиченої інформації [6] також виникають проблеми. Як вже було сказано, чим більша кількість навчальних зразків (z), та кількість ознак (n), тим ефективніше працює побудована мережа, але при цьому збільшується обсяг оброблюваної інформації, а отже збільшується необхідна кількість необхідної пам'яті, а також при цьому знижується швидкодія. А це означає, що пред'являються підвищені вимоги до комп'ютера. Ці вимоги стосуються оперативної пам'яті, ємкості твердого диска, частоти процесора. Наприклад, файл свопінгу займав приблизно 700 МБ. І це при тому, що сама програма для рішення задачі, яка була поставлена, займає відносно мало місця. Весь інсталяційний пакет займає приблизно 8 МБ і в нього входить сама інсталяційна програма з необхідними dll-файлами та іншими інсталяційними файлами, крім того, сюди входить необхідна база даних. Тому що для роботи даної програми необхідна наявність на комп'ютері Borland Database Engine, то у випадку відсутності цього пакету його необхідно встановити. Тому в інсталяційний пакет створеної програми входять файли, необхідні для установки Borland Database Engine. Для інсталяції необхідна наявність на твердому диску приблизно 27 МБ, з них для самої програми потрібно 19 МБ, та для Borland Database Engine – 8 МБ. При накопиченні великої кількості навчальних даних та при їх аналізі буде потрібно ще приблизно 15 МБ на робочому диску (крім обговорених раніше 700 МБ для файлу свопінгу).

При тестуванні мережі також виникають часові витрати. Так, наприклад, якщо тестувати мережу для всіх комбінацій по 5 користувачів з 8, за умови, що не проводиться усереднення навчальних даних ($usr=неправда$) та використовується по 1000 навчальних зразків для кожного класу ($z=1000$), тоді один такий експеримент займає приблизно 4 години. А якщо проводити експерименти, наприклад, для всіх комбінацій по 5 користувачів з 20, тоді час збільшується до декількох діб. Крім того, при проведенні експериментів, в яких для розрахунків вибирається оптимальне значення ширини функції активності (σ), тоді часові витрати збільшуються як мінімум у 20 разів.

Усі ці експерименти проводилися на комп'ютері Pentium III-500 з оперативною пам'яттю – 192 МБ, на якому була встановлена операційна система Windows 98. Потім експерименти були повторені на комп'ютері Athlon XP+2000 з оперативною пам'яттю 224 МБ, на якому була також встановлена операційна система Windows 98. При цьому час проведення кожного експерименту зменшився приблизно в 3–5 разів, завдяки чому вдалося провести більшу кількість експериментів.

Висновки. Таким чином, можна зробити висновок, що головними недоліками даного підходу є часові витрати для накопичення навчальних даних та їх обробки, а також підвищені вимоги до обчислювальної техніки, яка використовується.

Як указувалося вище, при проведенні тестування нейронної мережі навіть на сучасних ПЕВМ, час обчислень може досягати декількох діб. Для скорочення часових витрат пропонується застосування спеціалізованих

обчислювальних засобів із використанням програмованої логіки. Реконфігуруємі обчислювальні засоби на базі програмованих логічних інтегральних схем (ПЛІС) останнім часом знаходять все більш широке розповсюдження.

Відомо багато способів взаємодії програмної логіки з існуючою обчислювальною технікою. Оптимальним по співвідношенню вартості розробки і швидкості обміну даними між ПЛІС і центральним процесором є застосування так званих реконфігуруємих спеціалізованих обчислювачів (PCO). Дані пристрої представляють собою плату розширення, яку підключають до ПЕОМ за системною шиною [7]. Масове виробництво таких виробів веде до зниження їх вартості, як це відбулося з персональними комп'ютерами.

Головними труднощами при використанні таких пристроїв є розробка конфігурації – обчислювальної структури, що повинна завантажуватися в ПЛІС. Для зниження трудомісткості даної задачі запропоновано методику оптимізації алгоритмічного забезпечення PCO, що дозволяє максимально автоматизувати процес розробки конфігурації.

Список літератури: 1. *Каллан Р.* Основные концепции нейронных сетей.: Пер. с англ. – М.: Издательский дом “Вильямс”, 2001. – 290 с. 2. *Архангельский В.И., Богаенко И.Н., Грабовский Г.Г., Рюминин Н.Л.* Нейронные сети в системах автоматизации. – К.: “Техника”, 1999. – 364 с. 3. *Высоцкая Е.А., Давиденко А. Н.* Количественный и качественный анализ учебных данных с целью повышения эффективности аутентификации пользователей компьютерной системы при помощи нейронных сетей // Збірник наукових праць Інституту проблем моделювання в енергетиці. – К.: НАН України, 2003. – Вип. 24. – С. 110–116. 4. *Высоцкая Е.А.* Компьютерное моделирование задач аутентификации пользователя компьютерных систем с помощью вероятностных нейронных сетей. – К.: Інститут проблем моделювання в енергетиці НАН України. – 2004. – Вип. 24. – С. 3–9. 5. *Высоцкая Е.А., Давиденко А.Н.* Исследование эффективности применения вероятностных нейронных сетей для решения задачи аутентификации пользователя компьютерных систем / Науково-технічний збірник Національного технічного університету України “КПІ”, Міністерства освіти і науки України, Департаменту спеціальних телекомунікаційних систем та захисту інформації служби безпеки України “Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні”. – К.: 2004. 6. *Высоцкая Е. А., Давиденко А.Н.* Определение критических параметров при выборе биометрической системы аутентификации // Збірник наукових праць Інституту проблем моделювання в енергетиці. – К.: НАН України, 2004. – Вип. 28. 7. *Гильгурт С.Я.* Решение задач матричной арифметики с применением реконфигурируемых вычислений // Збірник наукових праць Інституту проблем моделювання в енергетиці. – К.: ПІМЕ НАН України, 2003. – Вип. 20. – С. 49–55.

Поступила в редакцію 11.10.2005