АВТЕНТИФІКАЦІЯ КОРИСТУВАЧА ЗА ДОПОМОГОЮ ЛІНГВІСТИЧНОГО МОДЕЛЮВАННЯ

УДК 811.93+004.042

ЄРМАКОВА Анастасія Анатоліївна
студентка кафедри ВТ НТУУ «КПІ»,
Наукові інтереси: лінгвістичне моделювання, приховані Марковські моделі.
nasti.carrot@gmail.com

БАКЛАН Ігор Всеволодович
канд.техн.наук, доцент кафедри АСОІУ НТУУ "КПІ",
Наукові інтереси: аналіз та прогнозування часових рядів, лінгвістичне моделювання,
приховані Марковські моделі, розпізнавання динамічних образів.
iaa@ukr.net

ВСТУП

Інформація, що зберігається на комп'ютері, є інтелектуальною власністю користувача, а тому потребує захисту, як і будь-яка інша власність. Залежно від можливих порушень у роботі системи та загроз несанкціонованого доступу до інформації використовуються різні види захисту інформації, які можна об'єднати у такі групи: морально-етичні, правові, адміністративні (організаційні), технічні (фізичні), програмні. Такий поділ є досить умовним. Зокрема, сучасні технології розвиваються в напрямку сполучення програмних та апаратних засобів захисту. Основою програмно-технічних засобів безпеки вважається ідентифікація і автентифікація. [1]

В наш час гостро постає питання захисту інформації та такої автентифікації користувача, яка унеможливлює несанкціонований доступ до даних. Технології невпинно розвиваються, потужності обчислювальних пристроїв зростають, отже, такі методи захисту, які раніше вважалися надійними, зараз вже не є такими.

Наприклад, до таких методів можна віднести логічні (введення паролів, ключових фраз з клавіатури). Щороку системи захисту потребують вводу все більш довгих паролів, більш надійних, бо короткі все легше підібрати.

Більш надійними методами автентифікації є ідентифікаційні, в яких носієм ключової інформації є фізичні об'єкти: магнітна карта, флеш-карта, штрих-кодова карта тощо. Недоліком цих методів є те, що карта може бути загублена або вкрадена зловмисником.

На мою думку, найнадійнішими методами автентифікації є біометричні. В їх основі полягає аналіз унікальних характеристик людини, наприклад: відбитки пальців, малюнок райдужної оболонки ока, голос, обличчя.

Отже, напрямком дослідження було обрано аналіз біометричного методу автентифікації за допомогою руху курсору, моделювання автентифікації за допомогою

лінгвістичного моделювання, порівняльний аналіз та оцінка цього методу.

Методи вирішення завдань дослідження та їх порівняльні оцінки. Існує багато методів, за допомогою яких можна вирішити завдання автентифікації користувача за рухом курсору. Серед них – нейронні мережі, байєсові мережі, метод групового урахування аргументів та інші.

Штучна нейронна мережа (ШНМ) - математична модель, побудована за принципом організації та функціонування біологічних нейронних мереж - мереж нервових клітин живого організму. Це поняття виникло при вивченні процесів, що протікають в мозку, і при спробі змоделювати ці процеси. Першою такою спробою були нейронні мережі У. Маккалок і У. Питтса. Після розробки алгоритмів навчання одержувані моделі стали використовувати в практичних цілях: в задачах прогнозування, для розпізнавання образів, в задачах управління та ін.

ШНМ є систему з'єднаних і взаємодіючих між собою простих процесорів (штучних нейронів). Такі процесори зазвичай досить прості (особливо в порівнянні з процесорами, використовуваними в персональних комп'ютерах). Кожен процесор подібної мережі має справу тільки з сигналами, які він періодично отримує, і сигналами, які він періодично посилає іншим процесорам. І, тим не менше, будучи з'єднаними в досить велику мережу з керованим взаємодією, такі локально прості процесори разом здатні виконувати досить складні завдання.

Нейронні мережі не програмуються в звичному сенсі цього слова, вони навчаються. Можливість навчання - одна з головних переваг нейронних мереж перед традиційними алгоритмами. Технічно навчання полягає в знаходженні коефіцієнтів зв'язків між нейронами. У процесі навчання нейронна мережа здатна виявляти складні залежності між вхідними даними і вихідними, а також виконувати узагальнення. Це означає, що в разі успішного навчання мережа зможе повернути вірний результат на підставі даних, які були відсутні в навчальній вибірці, а також неповних і / або «зашумлених», частково перекручених даних. [2]

Незважаючи на широкий спектр можливостей ШНМ, вирішення завдань з їх допомогою супроводжує ряд недоліків:

- більшість підходів для проектування ШНМ є наближеними і часто не призводять до однозначних рішень;
- для побудови моделі об'єкта на основі ШНС слід дотримуватися багатоциклового налаштування внутрішніх елементів і зв'язків між ними;
- проблеми, що виникають при підготовці навчальної вибірки, пов'язані з труднощами знаходження достатньої кількості навчальних прикладів;
- навчання мережі в ряді випадків призводить до тупикових ситуацій;
- тривалі часові витрати на виконання процедури навчання часто не дозволяють застосовувати ШНМ в системах реального часу;
- поведінка навченої ШНМ не завжди може бути однозначно передбаченою, що збільшує ризик застосування ШНМ для управління дорогими технічними об'єктами. [2]

Баєсова мережа - графічна імовірнісна модель, що представляє собою безліч змінних і їх імовірнісних залежностей по Баєсу. Наприклад, баєсова мережа може бути використана для обчислення ймовірності того, на що хворий пацієнт за наявністю або відсутністю ряду симптомів, ґрунтуючись на даних про залежність між симптомами і хворобами. Математичний апарат баєсових мереж створений амери-

канським ученим Джудой Перлом, лауреатом Премії Тьюринга (2011).

Формально, баєсова мережа - це спрямований ациклічний граф, кожній вершині якого відповідає випадкова змінна, а дуги графа кодують відносини умовної незалежності між цими змінними. Вершини можуть представляти змінні будь-яких типів, бути зваженими параметрами, прихованими змінними або гіпотезами. Існують ефективні методи, які використовуються для обчислень і навчання баєсових мереж. [3]

Недоліками баєсових мереж є:

– алгоритми кластеризації ефективні для розріджених мереж, але надзвичайно повільні для щільних мереж;
– алгоритми узгодження ефективні для розріджених мереж з невеликими циклічними розрізами, але мінімізація циклічних розрізів є NP-складною задачею;
– алгоритм реверсування ребра гарантує знаходження оптимального рішення реверсування дуг, але потребує неформального перевизначення ймовірностей подій в окремих вузлах;
– символічний алгоритм дає можливість працювати з параметрами у символічній формі без визначення їх числових оцінок, але потребує використання спеціальних програм для здійснення символьних обчислень, дуже неефективний при роботі з великими мережами та/або великою кількістю символьних параметрів;
– алгоритми стохастичного вибору ефективні у випадку відповідності вибіркового розподілу реальному спільному розподілу ймовірностей, але потребують багато часу на сходження;
– алгоритми спрощення моделі - обчислюваний час зменшується зі зменшенням станів вершини, зменшення станів вершини може суттєво впливати на точність результатів;
– алгоритм циклічного поширення для графів з циклами може давати слабкі результати або навіть не сходитися;
– алгоритми пошуку ефективні тільки для сильно асиметричних поширень, неефективні для складних багатозв'язних та багаторівневих мереж, що мають вершини з багатьма значеннями.[4]

Метод групового урахування аргументів (МГУА) - сімейство індуктивних алгоритмів для математичного моделювання мультипараметричних даних. Метод заснований на рекурсивному селективному відборі моделей, на основі яких будуються більш складні моделі. Точність моделювання на кожному наступному кроці рекурсії збільшується за рахунок ускладнення моделі.

Автор методу - академік НАНУ Олексій Григорович Івахненко. [5]

Метод запозичує ідеї з біології, а саме механізми еволюції:

1. схрещування або гібридизація батьківських пар (аргументів) і генерація нащадків;

2. селекція і відбір кращих.

Недоліки методу групового урахування аргументів:

– при близьких експериментальних точках можливо явище виродженість матриці нормальних рівнянь Гауса, внаслідок чого виникає необхідність застосування спеціальних методів регуляризації;

– метод дає точкову модель (прогнозу), а в ряді випадків бажано мати довірчий інтервал, який характеризує точність прогнозу. [6]

Аналіз робіт, присвячених порівнянню відомих методів ідентифікації та автентифікації, дозволив виявити ряд недоліків цих робіт: обмежене коло розглянутих

методів, відсутність чітко визначених показників оцінки їх якості, відсутність системності при проведенні оцінювання, відсутність, в більшості випадків, кількісних характеристик (оцінки виражені в нечіткій лінгвістичній формі), велика доля суб'єктивізму при оцінюванні зумовлена в тому числі комерційними (маркетинговими) інтересами. [1]

Загальна методика проведення досліджень. Лінгвістична модель — побудована на основі лінгвістичного моделювання сукупність символьних (лінгвістичних) послідовностей за обраними параметрами лінгвістизації та відновлена на її основі формальна граматика.

Лінгвістична модель динамічного процесу складається з наступних елементів

<D,I,L,G>

де D — сукупність часових рядів динамічного процесу та рядів, похідних від вхідних даних,

I — спосіб та правила інтервалізації,

L — морфізм відображення інтервального представлення ряду на певний алфавіт,

G — відновлена граматика динамічного процесу.

Великий внесок до галузі побудови лінгвістичних моделей та їх ймовірнісних гібридів внесли Марков А.А.(ст.), Канторович Л.В., Колмогоров А.М., Хомский Н., Апресян Ю.Д., Фу К.С., Рабінер Л. та ін.

Лінгвістичне моделювання — комплекс методів, методик та алгоритмів, які використовують процес перетворення числових масивів інформації до лінгвістичних послідовностей на основі яких відновлюється формальна граматика.

Лінгвістичне моделювання повинно забезпечувати:

- обґрунтований вибір інтервалів для виконання задач лінгвістизації (інтервалізація);

- ефективне перетворення числових масивів даних до лінгвістичних ланцюжків;

- підходів вивчення впливу обраних параметрів лінгвістизації на кінцеві результати застосування лінгвістичного моделювання;

- відновлення за лінгвістичними ланцюжками формальних граматик, в тому числі на ймовірнісних граматик та використання апарату прихованих марковських моделей для побудови гібридних моделей на основі лінгвістичного моделювання;

- інтеграцію лінгвістичних моделей з іншими обчислювальними парадигмами та створенні на їх основі гібридних процедур для вирішення різноманітних практичних завдань.

Лінгвістизація - процес перетворення часових рядів до сукупності лінгвістичних послідовностей, на основі яких будується формальна граматика.

Головним завданням лінгвістичного моделювання є перетворення чисельних рядів, експериментальних даних, багатомірних даних до лінгвістичних послідовностей та відновлення за ними формальної граматики мови відповідного характеру для вирішення наступного спектру проблем: аналіз та прогнозування числових рядів, розпізнавання образів різноманітної природи, автентифікація користувача за його рухами, розпізнавання емоційного стану оператора, діагностика хвороб опорно-рухового апарату операторів складних технічних систем на ранніх стадіях захворювання.

Лінгвістичне моделювання базується на трьох основних підходах: структурний підхід та математична лінгвістика, інтервальні обчислення та робастні методи, сучасні методи ймовірнісного моделювання. [7]

На основі лінгвістичного підходу можна будувати гібридні моделі. Основні принципи

гібридизації: принцип неоднорідності, принцип плюралізму, принцип системного аналізу неоднорідного завдання, принцип конструктора, принцип пріоритету знань, принцип поступовості, перший та другий принципи спадкування, принцип самоорганізації агрегованої моделі, принцип повноти, принцип зниження продуктивності агрегованої моделі.

Класичні гібридні системи комбінують аналогові та дискретні моделі. Агрегативні системи моделюють аналітико-статистичні закономірності поведінки складних систем. Методологія інтелектуальних експертних систем дозволяє перебороти недоліки символьного підходу за рахунок комбінування із традиційними інформаційними технологіями та технологіями штучного інтелекту. [7]

Прихована марківська модель (ПММ) — це статистична марківська модель, у якій система, що моделюється, розглядається як марківський процес із неспостережуваними (прихованими) станами. ПММ може бути представлено як найпростішу динамічну баєсову мережу. Математичний апарат для ПММ було розроблено Леонардом Баумом зі співробітниками. Він тісно пов'язаний з більш ранньою працею про оптимальну нелінійну проблему фільтрування Руслана Стратоновича, який першим описав послідовно-зворотній алгоритм.

У простіших марківських моделях (таких як ланцюги Маркова) стан є безпосередньо видимим спостерігачеві, і тому ймовірності переходу станів є єдиними параметрами. У прихованій марківській моделі стан не є видимим безпосередньо, але вихід, залежний від стану, видимим є. Кожен стан має ймовірнісний розподіл усіх можливих вихідних значень. Тому послідовність символів, згенерована ПММ, дає якусь інформацію про послідовність станів.

Приховані марківські моделі відомі в першу чергу завдяки їхньому застосуванню в розпізнаванні часових шаблонів, таких як розпізнавання мовлення, рукописного введення, жестів, морфологічної розмітки, мелодій для акомпонування, часткових розрядів та в біоінформатиці.

Приховані марковські моделі можуть розглядатися як узагальнення сумішевої моделі, де приховані змінні, що контролюють, яка складова суміші обиратиметься для кожного спостереження, пов'язані марковським процесом, а не є незалежними одна від одної. [8]

МЕТОДИ РОЗРАХУНКІВ, ГІПОТЕЗИ

Розглянемо динамічний метод біометричної автентифікації, який ґрунтується на особливостях руху користувача маніпулятором «миша» або пальцем чи стилусом по сенсорному екрану.

Для визначення "свій-чужий", користувачеві пропонується за допомогою маніпулятора "миша" зробити декілька рухів. Ці дії обробляються системою і записуються в вектор даних. Далі система розпізнавання порівнює характеристики отриманого вектору, з еталонними, які було отримано на етапі реєстрації користувача.

Така ідентифікація може провадитись не тільки під час первинного входу в систему, але і під час поточної роботи, що значно підвищує надійність захисту.

Дослідження буде проводитися за допомогою програмного забезпечення, яке буде знімати показання координат курсору через рівні проміжки часу. Після цього будуть вираховуватися дельти – різниці між попереднім та наступним показаннями координат X та Y. На основі масиву цих показників буде побудована лінгвістична модель, яка потім буде порівнюватись з наступними моделями з деяким припущенням після спроб автентифікації.

Існує гіпотеза, що автентифікація за допомогою лінгвістичного моделювання буде значно швидша за автентифікацію за допомогою інших методів.

ВИСНОВКИ

Всі методи, що використовувалися для автентифікації користувача за допомогою руху курсору, є досить повільними, що є неприпустимим для систем реального часу. Метод лінгвістичного моделювання, навпаки, дозволить прискорити процес авторизації. На даний час ще не було досліджено експериментально процес автентифікації за допомогою лінгвістичного моделювання, отже розробка є новою та революційною.

ЛІТЕРАТУРА:

1. Hadetska Z.M., Omelchuk D.G., Litvin R.V. Identyfikacija i autentyfikacija – metody zahystu vid nesankcionovanoho dostupu //Vostochno-Evropeiskyi zhurnal peredovyh tekhnologyi. – 2013. - №1(62). – S.8-10. - ISSN 1729-3774.
2. Tadeusevich Ryshard, Borovik Barbara, Gonchazh Tomash, Lepper Bartosh. Elementarnoe vvedenie v tehnologiiu neironnyh setei s primerami programm. / Perevod I.D. Rudinskogo. – M.: Goryachaia liniia – Telekom, 2011. – 408 s. — ISBN 978-5-9912-0163-6.
3. Korb Kevin B. Bayesian Artificial Intelligence. — CRC Press, 2004. — ISBN 1-58488-387-1. http://www.csse.monash.edu.au/bai/
4. Kuzmina N.F., Pietukh A.M. Ohliad metodiv obchyslennia baiesovykh merezh //Visnyk SumDU. Seriia "Tekhnichni nauky". – 2012. - №1. – S.112-117.
5. Madala H.R., Ivakhnenko A.G. Inductive Learning Algorithms for Complex Systems Modeling. - CRC Press, 1994 - 368p.- ISBN 0-8493-4438-7
6. Zaichenko Iu.P. Nechetkii metod induktivnogo modelirovaniia v zadachah prognozirovaniia makroekenomicheskih pokazateley //System. doslidzh. ta inform. tekhnologii. – 2003. - №3. – S.25-45.
7. Baklan I.V. Lingvistychne modeliuvannia: osnovy, metody, deiaki prykladni aspekty //Systemnyie tehnologii. – 2011. - №3(74). – S.10-19. – ISSN 1562-9945.
8. Satish L, Gururaj B.I. Use of hidden Markov models for partial discharge pattern classification //IEEE Transactions on Dielectrics and Electrical Insulation. – 2003. - №28. – S.172-182. - ISSN 0018-9367

Рецензент: д.т.н., проф. Ходаков В.Е.
Херсонский национальный технический университет