

8. Солодка Т. В. Контрольное тестирование как метод контроля результатов учебной деятельности студентов : дис. ... канд. пед. наук : 13.00.01 / Т.В.Солодка – Харьков, 1994. – 170 с.
9. Тестові технології оцінювання ключових і предметних компетентностей учнів основної і старшої школи: Монографія / За ред. Ляшенко О.І., Жука Ю.О. – К.: Педагогічна думка, 2014.– 200 с.
10. Чельшкова М. Б. Теория и практика конструирования педагогических тестов : учебное пособие / М. Б.Чельшкова – М. : Логос, 2002. – 432 с.

Тестовый контроль как средство оценки профессиональной компетентности будущих учителей информатики

Ткачук Г.В.

Аннотация. Описано проблему тестового контроля знаний в процессе оценивания профессиональных компетенций будущих учителей информатики. Определены организационные этапы подготовки для проведения тестового контроля, на которых предусматривается подготовка, разработка, апробация и внедрение разработанных тестов. Охарактеризованы особенности подготовки тестов по дисциплине «Технологии разработки веб-приложений» и представлены примеры тестовых заданий для оценивания предметных компетентностей. Предложено спецификацию итогового теста, в соответствии с которой определяется количество тестовых заданий по отдельным модулям и темам учебной дисциплины.

Ключевые слова: тестовый контроль, тестовое задание, методы тестирования, компетентности, спецификация.

Test control as a means evaluation of professional competence of future teacher of computer science

Tkachuk G. V.

Resume. Described problem of test control in process the professional competence of future teachers of computer science. Identified organizational stages of preparation for test control: training, development, testing and implementation of the developed test. Characterized the features of preparation of tests in the discipline "Technology of development of web-application" and presented examples of organizing test tasks for the evaluation of the subject competence. Proposed a specification of a final test, which distributes and determines the number of tasks for modules and themes discipline.

Keywords: test control, test task, test methods, competence, specification.

УДК 378.1 : 004.89

Іваськів І. С.
ПрАТ Газінтек

Биометрична ідентифікація користувачів систем дистанційного навчання на основі методів виявлення аномалій засобами штучних нейронних мереж

Анотація. В статті розглянуто питання біометричної ідентифікації користувачів в системах дистанційного навчання, електронних курсів, віртуальних лабораторій. Обґрунтовується можливість застосування детекторів аномалій на базі штучних нейронних мереж для аналізу динаміки клавіатурного ритму користувачів з метою ідентифікації у системі. Предметом дослідження є методика використання штучних нейронних мереж в задачах біометричної ідентифікації. Метою дослідження є оптимізація архітектури та параметрів штучної нейронної мережі в режимі детектора аномалій для задач ідентифікації користувачів систем дистанційного навчання.

Ключові слова: системи дистанційного навчання, виявлення нетипової поведінки, виявлення аномалій, штучні нейромережі, біометрична ідентифікація, аналіз клавіатурного ритму.

Спираючись на теоретичні дослідження багатьох вчених, зокрема В.П. Андрущенко, В.Ю. Бикова, М.І. Жалдака, В.Г. Кінельова, К.К. Коліна, В.Г. Кременя, Н.В. Морзе, С.А. Ракова, Ю.С. Рамського, С.О. Семерікова, О.М. Спіріна, Ю.В. Триуса, серед важливих тенденцій в розвитку освіти в інформаційному суспільстві можна виокремити створення інформаційного освітнього середовища на базі досягнень науково-технічного прогресу в галузі інформаційно-комунікаційних технологій. Зокрема, В.Ю. Биков, М. П. Шишкіна, О. М. Спірін, Ю. Г. Запорожченко, В.П. Олексюк, досліджуючи проблематику інформатизації освіти, зазначають, що розвиток технологій хмарних обчислень, сервісів адаптивних інформаційно-комунікаційних мереж, засобів віртуального і мобільного навчання є важливим кроком на шляху розв'язання проблем доступності і якості навчання, що змінює уявлення про інфраструктуру організації процесу навчання.

Одним з напрямків удосконалення таких сервісів є запровадження якісних та зручних систем ідентифікації учнів в віддаленому віртуальному середовищі. Якщо система автентифікації дозволяє забезпечити допуск чи не допуск користувача в середовище, система ідентифікації дозволяє стверджувати про, наприклад, самостійність виконання завдань певним користувачем.

Биометричні технології ідентифікації зазвичай поділяють на фізіологічні (відбиток пальців, обличчя, сітківки ока) та поведінкові (підпис, клавіатурна динаміка). Разом з поширенням інтернет-

технологій у різні галузі діяльності та необхідності віддаленої ідентифікації користувачів спостерігається значне зростання інтересу до поведінкової біометрики. Це є наслідком простоти використання сервісів кінцевим користувачем, прозорості та великої кількості потенційних способів застосування.

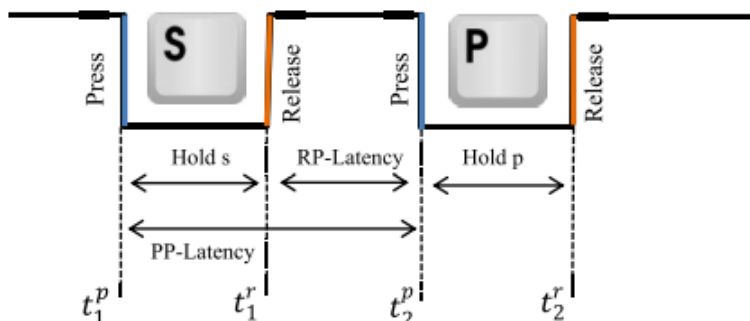
Біометрика клавіатурного введення (keystroke biometrics, keystroke dynamics) – метод автоматичної автентифікації/ідентифікації, що базується на розпізнаванні патернів (ритму) клавіатурного набору. На відміну від інших біометричних систем, що можуть базуватись на дорогому обладнанні, аналіз клавіатурного введення не потребує значних вкладень.

Перші дослідження з використання набутих патернів в клавіатурному введенні були проведені Гейнсом ще у 80 роках минулого століття [1]. Експеримент проводився з невеликою вибіркою даних, отриманих внаслідок клавіатурного введення семи секретарів. Тестування на статистичну незалежність їх профілів було проведено з використанням системи *T-Test* з гіпотезою, що середні значення часу диграф в різних сесіях однакові але з різною дисперсією. Подібні дослідження проводились Леджетом [2] за участю 17-ти програмістів, проте з іншим підходом до ідентифікації. Дослідником використовувався так званий метод безперервної ідентифікації, коли аналіз клавіатурного введення проводився на протязі всієї робочої сесії.

Більшість сучасних алгоритмів ідентифікації за допомогою аналізу клавіатурного введення базується на методах виявлення нетипової поведінки, аномалій. Виявлення нетипової поведінки в системі є задачею знаходження патернів (шаблонів) в даних, які не відповідають моделі «нормальної», очікуваної поведінки. Нетипова поведінка системи в літературі з аналізу даних та статистики також згадується як аномальність, відхилення, викид (outlier). Hawkins D. в [3] формально визначає концепцію викиду так: «Викид – це спостереження, яке відхиляється настільки від інших спостережень, що виникають підозри, що він був породжений іншим механізмом».

В більшості інформаційних систем дані створюються в одному або більше генеративних процесів, і за ними можна відображати активність в системі або спостереження за сутностями. Коли генеративний процес перебігає «незвичним способом», це є результатом створення викидів чи аномалій. У цих програмних застосунках дані пов'язуються «нормальною» моделлю і аномалії розпізнаються як відхилення від цієї нормальної моделі. Це досить просто виконати для систем, в яких поведінка може бути визначена за допомогою простих математичних моделей – наприклад, систем які характеризуються розподілом Гауса з відомим середнім і стандартним відхиленням. Проте, найбільш цікаві системи реального світу складно змінюються в часі. Для вивчення характеристик таких систем на основі спостережуваних даних стають у нагоді методи машинного навчання та штучні нейронні мережі.

В загальному випадку біометрична ідентифікація на основі аналізу клавіатурного ритму включає 2 фази. Перша фаза – етап збирання даних та побудови моделі. За попереднього кількаразового введення користувачем певної паролльної фрази фіксується набір ознак – час між послідовними натискуваннями окремих клавіш, час утримання клавіш (Рис. 1). Також можуть фіксуватись порядок використання спеціальних клавіш, спосіб виправлення помилок. На основі цих ознак будується певна модель. На етапі тестування, під час введення паролльної фрази визначається певний ступінь нормальності чи аномальності отриманих даних згідно розробленої попередньо моделі. На основі цього показника робиться припущення про те, чи дане тестове введення було проведено тією самою особою, на даних клавіатурної динаміки якої розроблялася попередня модель, чи зловмисником, в разі спроби підміни собою оригінального користувача.



- Press-Press (Keydown - Keydown) – час між послідовними натискуваннями;
- Release-Press (KeyUp - Keydown) – час між відпусканням клавіші та натискуванням наступної
- Hold – час утримування клавіші.

Рис. 1. Найбільш популярні ознаки, які застосовуються в моделях динаміки клавіатурного введення

Протягом останнього десятиріччя запропоновано велику кількість алгоритмів – моделей даних для автентифікації/ідентифікації на основі аналізу клавіатурного введення. Проведемо короткий огляд популярних моделей.

Euclidean

Класичний алгоритм виявлення аномалій. Кожне клавіатурне введення пароля моделюється точкою в p -вимірному просторі, де p – кількість ознак в часовому векторі. Модельні дані

розглядаються як хмара точок, і обчислюється ступінь аномальності тестового вектора, базуючись на його віддаленості від центра цієї хмари. Формально кажучи, під час підготовчої фази визначається середній вектор часових векторів вибірки. В тестовій фазі ступінь аномальності визначається як квадрат евклідової відстані між тестовим вектором та середнім вектором.

Euclidean (normed)

Нормалізований бінарний класифікатор на базі метрики мінімальної відстані. В підготовчій фазі середній вектор визначається як в стандартному евклідовому детекторі. Ступінь аномальності визначають «нормалізуючи» квадрат евклідової відстані між тестовим та середнім вектором шляхом ділення на добуток норм цих двох векторів:

Якщо x – середній вектор, y – тестовий вектор, то ступінь аномальності:

$$score = \frac{d}{\|x\|\|y\|}$$

Manhattan

Модель подібна до стандартного евклідового детектора, проте замість евклідової метрики відстані використовується *манхеттенівська відстань* (відстань міських кварталів).

Manhattan (filtered)

В підготовчій фазі обчислюється середній вектор і стандартне відхилення кожної ознаки. Будь який часовий вектор, який різниться на більш ніж три стандартні відхилення від середнього, видаляється з вибірки і обчислюється новий середній вектор, вже без врахування цих екстремальних ознак. В тестовій фазі манхеттенівська відстань від цього нового вектора є мірою аномальності.

Manhattan (scaled)

Міра аномальності обчислюється подібно до манхеттенівської, проте кожен вимір масштабується на середню величину абсолютного відхилення кожної ознаки від середнього значення:

$$\sum_{i=1}^p \frac{|x_i - y_i|}{a_i}, \text{ тут } a_i - \text{ середнє абсолютне відхилення ознаки}$$

Mahalanobis

Міра відстані у цій моделі більш складна за евклідову. Враховується кореляція поміж ознаками. Під час підготовчої фази обчислюються середній вектор та коваріаційна матриця часових векторів. В тестовій фазі міру аномальності обчислюють як відстань Махаланобіса між середнім та тестовими векторами:

$$(x - y)^T S^{-1} (x - y)$$

Найближчих сусідів (Nearest-neighbor)

У цій моделі міру аномальності обчислюють як відстань Махаланобіса між тестовим вектором та найближчим вектором з підготовчої вибірки.

Підрахунок викидів (Outlier-counting) – z-score

Статистичний метод. Під час підготовчої фази обчислюється середнє значення та стандартне відхилення кожної часової ознаки. В тестовій фазі обчислюється абсолютне значення z-score кожної ознаки за формулою:

$$|x_i - y_i|/s_i, \quad (3.4)$$

де x_i, y_i – i -ті ознаки середнього та тестового векторів відповідно, s_i – стандартне відхилення від підготовчої фази. Міра аномальності розраховуються як кількість z-score, що перевищили певний пороговий рівень.

Нейронна мережа (стандартна)

Детектор описаний Хайдером [4] і побудований на базі штучної нейронної мережі прямого поширення з «навчанням» за допомогою алгоритму бекпропагації (зворотнього поширення помилки). В мережі міститься p вхідних нейронів (за кількістю ознак – довжиною часового вектора), один вихідний нейрон та $2p/3$ прихованих нейрони. Детектор «натренований» видавати 1 на кожному навчаючому прикладі (мережа «навчається» лише на «нормальних» даних, тобто, даних оригінального користувача). Протягом тестової фази ступінь аномальності обчислюється як $1-s$, де s – величина, отримана на виході мережі. Якщо s наближається до 1.0 – тестовий вектор «подібний» до векторів з «навчаючого» набору.

Нейронна мережа (адаптивний фільтр)

Розроблена та протестована нами система побудована на базі штучної нейронної мережі, проте на відміну від описаної вище моделі мережа налаштована в режимі адаптивного фільтра, описаного нижче.

Пригадаємо формальне означення задачі класифікації, як одного з розділів машинного «навчання». Нехай X – множина описів об'єктів, Y – множина номерів (або міток класів). Існує невідома цільова залежність – відображення: $y^* : X \rightarrow Y$, значення якої відомі лише на об'єктах навчаючої вибірки $X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}$. Необхідно побудувати алгоритм $\alpha : X \rightarrow Y$, за допомогою якого можна класифікувати довільний об'єкт $x \in X$. Якщо множина Y складається лише з двох елементів, то має місце бінарна (двокласова) класифікація. В цьому випадку умовно вважають, що $Y = \{-1, 1\}$ і один з класів – позитивним, другий негативним.

Розглянемо випадок бінарної класифікації, коли структуру множини X^m можна подати у вигляді $\{(x_1, y_1), \dots, (x_m, y_1)\}$, тобто відома тільки така звана *позитивна* вибірка. Ціль залишається незмінною. В цьому випадку має місце *однокласова* класифікація, або *класифікація з «навчанням» на основі тільки позитивних прикладів*.

Одним із способів реалізації однокласового класифікатора є використання багатоповірного перцептрона (штучної нейронної мережі прямого поширення) в якості адаптивного фільтра. «Навчання» нейронних мереж на основі тільки позитивних прикладів було розглянуто в роботах [5], [6].

Нехай існує набір векторів V_n , кожен з яких є *позитивним* розмірності m . Виберемо деяку метрику M , за якою буде описуватися відстань між векторами (в якості метрики можна взяти відстань евклідову чи Чебишова). Побудуємо штучну нейронну мережу прямого поширення з m вхідними нейронами, h нейронами прихованого шару та m вихідними нейронами. Крім того, прихований шар має сигмоїдну функцію активації, вихідний – лінійну.

Для «навчання» мережі за методом зворотного поширення помилки будемо використовувати «навчаючу» вибірку X^n виду: $\{(x_1, x_1), \dots, (x_n, x_n)\}$ (для кожного прикладу в якості вектора результатів використано той самий вектор). Тобто побудована штучна нейронна мережа буде функціонувати як *адаптивний фільтр*, через який поданий на вхід сигнал (вектор) повинен без спотворення подаватися на вихід. Отже, для векторів, «подібних» до векторів «навчаючої» вибірки, відстань $M(x_i, y_i) \rightarrow 0$, де y_i – вихідний вектор ШНМ за подавання на вхід вектора x_i . Після «навчання» мережі, необхідно пропустити всю «навчаючу» вибірку через ШНМ і отримати порогове значення метрики $threshold = \max_{\forall i \in (1, n)} M(x_i, y_i)$. Після цього для отримання висновку чи є довільний вектор позитивним, достатньо визначити засобом побудованої та «навченої» ШНМ вектор y та перевірити – чи не перевищує значення $M(x, y)$ значення $threshold$.

Схему ШНМ, на основі якої реалізується функціонування детектора аномалій, подано на рис. 2. Ефективність роботи такого детектора аномалій залежить від наступних параметрів, які будуть експериментально встановлені в ході дослідження:

- h – кількість нейронів прихованого шару, від чого залежить обсяг «пам'яті» ШНМ; якщо h буде більше або рівне m , то виникає ризик «перенавчання»;
- lr – коефіцієнт швидкості «навчання» (lr , learning rate), від чого залежить швидкість «навчання», так і ефективність процесу «сходження»;
- α – параметр регуляризації (регуляризація типу L2) – його використання допомагає уникати «перенавчання» шляхом штрафування нейро-коефіцієнтів з великими значеннями.

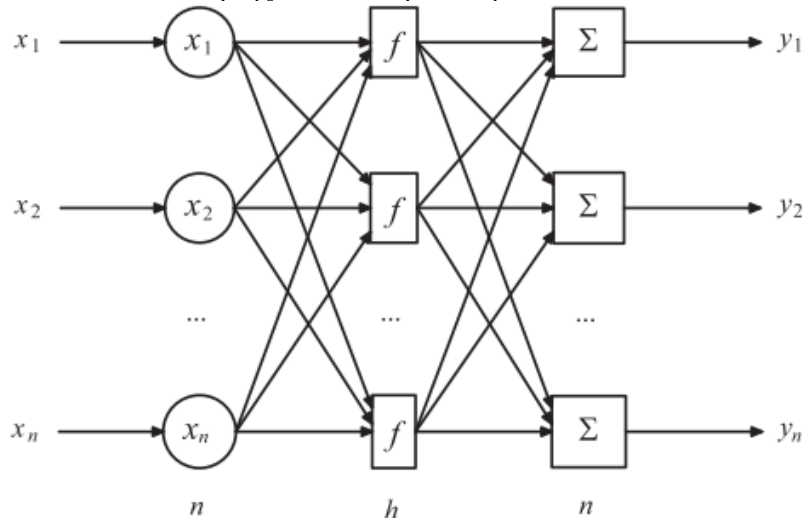


Рис. 2 Архітектура штучної нейронної мережі для задач виявлення аномалій

Дослідження описаної моделі проводилось з використанням мови програмування Python, бібліотеки машинного «навчання» Scikit-learn, в середовищі Jupyter Notebook. Для вимірювання ефективності моделей використовувались ROC – криві (Receiver Operating Characteristic – робоча характеристика приймача). Для побудови ROC-кривої використовують рейтинги «влучання» – True Positive Rate – оцінка якості визначення користувача як зловмисника) та False Positive Rate – оцінка якості помилкового визначення «нормального» користувача як зловмисника. Проводилось обчислення даних рейтингів за різних рівнів $threshold$ – порогу, та будувались відповідні графіки. ROC крива – розповсюджений метод оцінювання точності детекторів, його використання дозволяє проводити різні оцінювання такі як, наприклад, співвідношення хибних спрацювань та пропусків.

Однією з оцінок якості алгоритму класифікації є площа під даною кривою. Чим вона ближча до 1, тим краща якість моделі. Оптимізована в процесі даного дослідження структура та параметри штучної нейронної мережі дозволили досягнути метрики «площа під ROC-кривою – 0.948. Для досліджуваного набору з 30 вхідними ознаками це мережа прямого поширення з одним прихованим шаром, з 22-ма нейронами в ньому, функцією активації ReLU (усічене лінійне перетворення), вихідного шару з кількістю нейронів рівною кількості вхідних нейронів та без функції активації в ньому.

Отриманий графік оцінювання якості за методом кривої ROC подано на рис. 3.

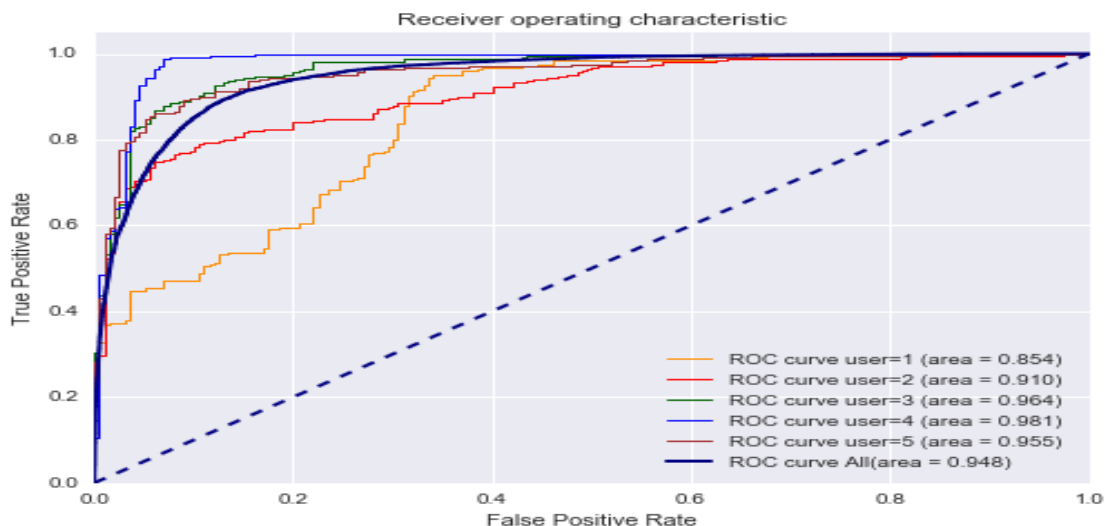


Рис. 3. Крива ROC (робочої характеристики приймача) побудованої моделі для 5-ти користувачів

Система біометричної автентифікації порівнювалась з системами, описаними в дослідженні [7]. В процесі даного дослідження розроблено уніфіковану методику тестування якості функціонування алгоритмів автентифікації на базі аналізу динаміки клавіатурного введення. В якості паролльної фрази було вибрано стрічку, відповідно до вимог надійності паролів: .tie5Roanl. В дослідженні брали участь 50 людей різного віку. Кожен з них повторив введення паролю 400 разів, під час поділених на кілька днів сесій.

Для порівняння якості роботи розробленої і «навченої» штучної нейронної мережі з іншими алгоритмами використані оцінки *equal-error rate* та *zero-miss false-alarm rate*. Для обчислення *equal-error rate* поріг вибраний таким чином, щоб рейтинг пропусків та помилкових спрацювань були рівними. Для обчислення *zero-miss false-alarm rate* поріг вибраний так, щоб кількість хибних спрацювань була мінімальною за кількості пропусків, рівної нулю.

Результати порівняння різних алгоритмів подано в таблиці 1. Як видно з таблиці, під час використання детектора у режимі 100 відсоткового виявлення паролів, введених зловмисником, частка хибних спрацювань (пароль введений оригінальним користувачем, проте відхилений системою) становить близько 48%. Це становить незручність для користувачів, змушуючи їх до повторного введення пароля. Тому досліджувана система потребує подальшого доопрацювання та вдосконалення.

Таблиця 1

Середні значення Equal Error Rate та Zero-miss Rate

Детектор	Equal Error Rate	Zero-miss Rate
Manhattan (scaled)	0.096	0.468
Nearest Neighbor	0.100	0.468
Outlier Count (z-score)	0.102	0.782
Досліджувана система	0.102	0.483
SVM (one-class)	0.102	0.504
Mahalanobis	0.110	0.482
Mahalanobis (normed)	0.110	0.484
Manhattan (filter)	0.136	0.757
Manhattan	0.153	0.843
Euclidean	0.171	0.875
Euclidean (normed)	0.215	0.911
k-Means	0.372	0.989
Neural Network (standard)	0.828	1.000

Експеримент показав, що збільшення кількості ознак (в даному випадку – довжини пароля) позитивно впливає на якість функціонування детектора. Тому подальші дослідження доцільно проводити у цьому напрямку.

В проведеному дослідженні запропоновано методику використання штучної нейронної мережі для біометричної ідентифікації користувачів. Запроектовано та виконано програмну реалізацію системи біометричної ідентифікації на основі аналізу аномалій клавіатурного ритму оптимізованою нейронною мережею. Отримані результати еквівалентної частки помилок (рівної частки хибних спрацювань та хибних допусків в систему) - 0.102. В режимі детекції паролів, введених зловмисником, близької 100 відсоткам, отримується 48.3% хибних спрацювань. Хоча цей результат є одним з найкращих серед систем, які функціонують з використанням інших моделей, її практичне застосування можливе лише в якості допоміжного засобу перевірки користувачів, для допуску до

програмних систем чи онлайн ресурсів, як наприклад, системи дистанційного навчання (перевірка самостійності виконаного завдання користувачем).

Список використаних джерел

1. Gaines R. Authentication by keystroke timing: some preliminary results / Gaines R., Lisowski W., Press S. // Technical Report R-2526-NSF. – Santa Monica: RAND Corporation, 1980. – 51 p.
2. Leggett J. Verifying identity via keystroke characteristics / Leggett J., Williams G. // International Journal of Man-Machine Studies. – London: Academic Press Ltd, 1988. – pp.67-76.
3. Hawkins D. Identification of Outliers / Hawkins D. – Springer Netherlands, 1980. – 188 p.
4. Haider S. A multi-technique approach for user identification through keystroke dynamics / Haider S., Abbas A., Zaidi A.K. // IEEE International Conference on Systems, Man and Cybernetics, 2000. – pp. 1336–1341.
5. Manevitz L. Document Classification on Neural Networks Using Only Positive Examples / Manevitz L., Malik Y. // ACM SIGIR conference, 2000.
6. Большев А.К. Применение нейронных сетей для обнаружения вторжений в компьютерные сети / Большев А.К., Яновский В.В. // Вестник Санкт-Петербургского университета, Вып. 4. СПб., 2009. – С. 38-44.
7. Kevin S. Comparing Anomaly-Detection Algorithms for Keystroke Dynamics / Kevin S. // IEEE, 2009. – pp.125-134.

Биометрическая идентификация пользователей систем дистанционного образования на основании методов обнаружения аномалий средствами искусственных нейронных сетей

Иваськив И.С.

Аннотация. В статье рассмотрен вопрос биометрической идентификации пользователей систем дистанционного образования на основании анализа их клавиатурного ритма с использованием средств искусственной нейронной сети. Описываются основные алгоритмы и модели данных, применяемые в задачах моделирования клавиатурной динамики пользователей. Описываются структура и параметры сети – детектора аномалий, применяемой для биометрической идентификации пользователей системы дистанционного обучения.

Ключевые слова: системы дистанционного образования, обнаружение аномалий, биометрическая идентификация, анализ клавиатурного ритма.

Biometric users identification in online education systems based on anomaly detection methods by means of artificial neural networks

I.S. Ivaskiv

Resume. The article deals with the issue of biometric identification of users of distance education systems based on the analysis of their keyboard rhythm by means of an artificial neural network. The main algorithms and data models used in modeling problems of user's keyboard dynamics are described. The structure and parameters of the artificial network – an anomaly detector used for biometric identification of users, are described.

Keywords: online learning, anomaly detection, biometric identification, neural networks.

УДК 37 004(07)

Ящик О. Б.

Тернопільський національний педагогічний університет імені Володимира Гнатюка

Зміцнення глобальної культури кібербезпеки в мережі Інтернет

Анотація. У статті досліджуються проблеми кібербезпеки в мережі Інтернет; вивчено історію її виникнення та розвитку; з'ясовані основні принципи вдосконалення глобальної культури безпечної мережевої взаємодії. Розглянуті питання, пов'язані із забезпеченням інформаційної безпеки особистості учнів в контексті професійної підготовки компетентних педагогів, здатних сприяти створенню інфобезпечного середовища в школі, а також навчити школярів захищатись від небезпечного і шкідливого інформаційного контенту.

Ключові слова: кібербезпека, мережева взаємодія, захист персональних інформаційних ресурсів, шкідливий інформаційний контент, інфо-безпечне навчальне середовище.

Актуальність проблеми. XXI століття характеризується інтенсивним використанням інформаційних технологій у житті суспільства. Широке проникнення соціальних мереж в повсякденне життя, використання Інтернету як основного джерела різноманітних відомостей внесли серйозні зміни в соціальні стосунки у світі. Сьогодні використання інформаційних технологій дає можливість людям з різних країн і континентів обговорювати актуальні проблеми в режимі реального часу, отримувати відомості безпосередньо з місця подій.

Усі великі світові ЗМІ мають представництва в мережі Інтернет, даючи нам унікальну можливість отримувати відомості з розмаїтих джерел, формувати загальне уявлення про те, що відбувається в освіті, бачити через таке спілкування відмінність у культурах (Рис. 1). Помітною особливістю Інтернету до останнього часу була можливість залишитися анонімним, приховати своє