УДК 004.4.056. **Пархоменко І. І., Молодан Б. О., Нечипорук В.В.**

ЗАСТОСУВАННЯ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ В СИСТЕМАХ АНТИВІРУСНОГО ЗАХИСТУ

В статті розглянуті основні методи штучного інтелекту в контексті застосування системами антивірусного захисту. Наведена оцінка їх ефективності та перспективи застосування. Розглянуто можливості створення “розумних” систем антивірусного захисту на основі штучного інтелекту.

Ключові слова: антивірус, комп’ютерний вірус, штучний інтелект, теорема Баєса, штучні нейронні мережі, метод опорних векторів.

Вступ. З метою забезпечення антивірусного захисту комп'ютерних систем були створені системи, які можуть класифікувати поведінку різних програм. У разі збігу з ситуацією, визначеною експертом, такі системи пропонують користувачеві припинити дії можливо шкідливого програмного забезпечення (ПЗ) і відновити зміни у системі. Однак більшість сучасних систем антивірусної безпеки не мають можливості самонавчання і оперують тільки закладеними в них правилами. Часта поява небажаного ПЗ, що використовує нові уразливості операційних систем, підвищує вимоги до систем антивірусної безпеки. Застосування методів штучного інтелекту (ШІ) дозволяє додати в системи захисту властивість самонавчання і забезпечити виявлення загроз «на льоту». У даній статті наводиться огляд методів ШІ, що можуть бути використані для захисту комп'ютерних систем.

Аналіз останніх досліджень та постановка проблеми. Питання ШІ та методів його створення розглянуті в джерелах[1,3], в дослідженнях не розглядаються використання ШІ в аспекті антивірусного захисту. *Актуальність* дослідження антивірусної безпеки невпинно росте, з постійним збільшенням кількості шкідливого ПЗ. *Метою* даної роботи є огляд методів ШІ в аспекті антивірусного захисту.

Основна частина. *Теорема Баєса* Одним з підходів до обробки не повністю певної інформації є ймовірнісне моделювання предметної області. У цій сфері широке поширення набули системи, засновані на теоремі Баєса (Bayes' Theorem). Теорема виражається формулою Баєса:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)},$$

де: P (H|X) – ймовірність гіпотези H при настанні причини X; P (X|H) - ймовірність присутності причини X при істинності гіпотези H; P (H) – апріорна ймовірність гіпотези H; P (X) - ймовірність настання причини X.

Ця проста формула лежить в основі багатьох сучасних систем штучного інтелекту, призначених для роботи в умовах невизначеності [1, 2]. Такі системи дають вірогідну оцінку, тому зазвичай не замінюють експерта, а надають йому допомогу в прийнятті рішення.

Розглянемо приклад спам-фільтра на основі теореми Баєса. При навчанні фільтра масив електронних листів ділиться на два класи: спам і корисна кореспонденція. Для кожного слова обчислюється частота його зустрічі в обох класах листів.

Позначимо F_S (W_i) - кількість спам-листів, в яких зустрілося слово W_i, а F_{NS} (W_i) - кількість корисних листів, в яких зустрілося слово W_i. У задачі присутні дві гіпотези: H_S - лист є спамом, H_{NS} – лист має корисну інформацію. Тоді ймовірність того, що поява слова W_i в листі означає спам, обчислюється за формулою:

$$P(W_i|H_S) = \frac{F_S(W_i)}{F_S(W_i) + F_{NS}(W_i)},$$

а ймовірність того, що слово W_i не вказує на спам в листі:

$$P(W_i|H_{NS}) = \frac{F_{NS}(W_i)}{F_S(W_i) + F_{NS}(W_i)}.$$

Вектор W включає всі слова нового листа. Тоді для нового листа ймовірність того, що воно спам, обчислюється за формулою Баєса наступним чином:

$$P(H_S|W) = \frac{P(W|H_S)P(H_S)}{P(W|H_S)P(H_S) + P(W|H_{NS})P(H_{NS})}.$$

Вважаючи апріорні ймовірності обох гіпотез однаковими, отримуємо:

$$P(H_S|W) = \frac{\prod_{j=1}^{m} P(W_j|H_S)}{\prod_{j=1}^{m} P(W_j|H_S) + \prod_{j=1}^{m} P(W_j|H_{NS})}.$$

Віднесення листа до спаму або до корисних листів проводиться звичайно з урахуванням заданого користувачем порога, значення якого складають 0,6 ÷ 0,8. Після прийняття рішення по листу, в базі даних оновлюються ймовірності для вхідних слів. Розглянутий метод простий в реалізації, ефективний (після навчання на досить великій вибірці листів відсікає до 95-97% спаму) та володіє можливістю навчання. Зазначені характеристики пояснюють той факт, що на основі теореми Баєса побудовано безліч сучасних спам-фільтрів. Для обходу традиційних спам-фільтрів спамери почали вкладати рекламну інформацію в зображенні, а текст в листі або відсутній, або не несе сенсу. Проти цього доводиться користуватися або засобами розпізнавання тексту, або старими методами фільтрації - «чорні списки» і регулярні вирази (так як такі листи часто мають стереотипну форму). Лабораторія Касперського в своїх продуктах реалізувала технологію розпізнавання тексту на вкладених картинках і пересилала на спам-фільтр. Розвитком імовірнісного підходу на основі теореми Баєса є Баєсовські мережі (Bayesian networks). Баєсова мережа являє собою модель, що відображає імовірнісні та причинно-наслідкові відносини між змінними і дозволяє скласти наочний опис повного спільного розподілу ймовірностей [2]. За структурою мережа є орієнтованим графом, в якому кожна вершина має деякі значення ймовірностей.

Для отримання працездатною баєсівської мережі її навчають на наборі даних, підготовленому експертами. При навчанні намагаються мінімізувати ризик виникнення помилки при роботі мережі в подальшому. Для цього використовуються спеціальні алгоритми, такі як градієнтний спуск, алгоритм EM (Expectation - Maximization, очікування - максимізація) та інші [2, 3].

Штучні нейронні мережі. Штучна нейронна мережа (НС) є спрощеною моделлю мозку і представляє набір нейронів, з'єднаних між собою певним чином [2, 4]. Нейронні мережі дозволяють вирішувати різні практичні завдання, пов'язані, в основному, з розпізнаванням і класифікацією образів. Безперечними перевагами НС є те, що вони можуть автоматично набувати знання в процесі навчання і мають здатність до узагальнення.

Основним елементом мережі є штучний нейрон (рис. 1) - математична модель біологічної нервової клітини.

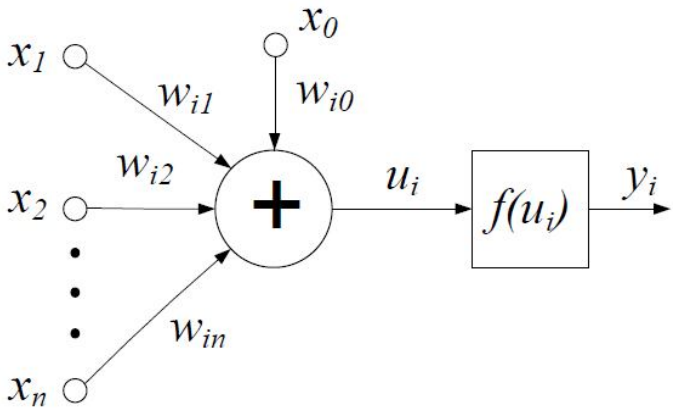

Рис. 1 Модель нейрона

У цій моделі вхідні сигнали x_j (j = 1, 2, ..., n) підсумовуються в і-м нейроні з урахуванням відповідних вагових коефіцієнтів w_{ij}. Також входить в суму x_0 – поріг (bias, сигнал поляризації), що визначає зменшення або збільшення вхідного сигналу на задану величину.

Сума надходить на вхід функціонального блоку $f(u_i)$, вихід якого є вихідним сигналом нейрона. Таким чином, роботу і-го нейрона можна описати наступній функцією:

$$y_i = f\left(\sum_{j=1}^{n} w_{ij}x_j + w_{i0}x_0\right).$$

Виходячи з виду функції $f(u_i)$, званою функцією активації, розрізняють кілька типів нейронів. Найбільш часто використовуваний – сигмоідальний нейрон, його функція активації має наступний вигляд:

$$f(u) = \frac{1}{1 + e^{-\beta u}}.$$

Окремі нейрони об'єднуються в мережі з різноманітною архітектурою. Нині широко застосовуються багатошарові мережі прямого поширення [5]. У цих мережах виходи нейронів одного прошарку служать входами для наступного прошарку (рис. 2)

Застосування НС для вирішення будь-якої задачі включає два етапи: етап навчання та етап розпізнавання. На етапі навчання на вхід НС подається навчальна вибірка, складається з заздалегідь відібраних і підготовлених вхідних і вихідних векторів. В відповідно до обраного алгоритмом навчання (наприклад, метод зворотного поширення помилки або метод сполучених градієнтів) відбувається налаштування вагових коефіцієнтів, в результаті якого при подачі на вхід НС навчального вектора на виході з'являється заданий вихідний вектор, що позначає клас вхідного вектора.

На етапі розпізнавання на НС надходить заздалегідь невідомий вхідний вектор. При цьому на виході з'являється вектор - результат розпізнавання, відповідно до якого вхідний вектор зараховується до одного з відомих класів.

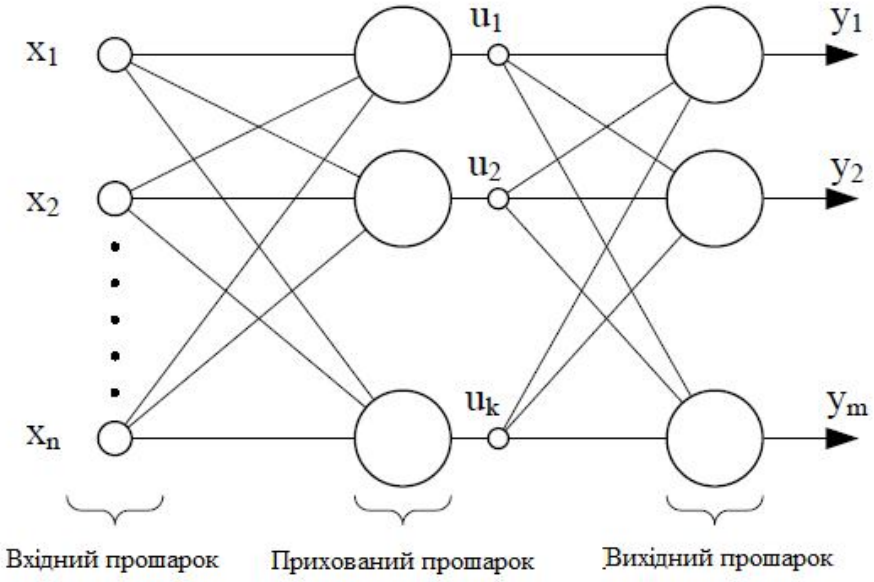

Рис. 2 Двошарова нейрона мережа

Таким чином, у разі використання НС у сфері антивірусної безпеки, будь-яка дія користувача або програми повинно бути представлено у вигляді вектора ознак, які подаються на вхід НС. В результаті проходження сигналів по мережі, на виході виходить вектор, що визначає, чи є дія шкідливою. Розглянемо застосування НС на невеликій практичній задачі з області інформаційної безпеки (приклад взято з [5]). За допомогою НС необхідно детектувати доступ до бази даних з боку якогось ПЗ, відмінного від автоматизованого робочого місця (АРМ) користувача, або визначити аномалії в роботі користувача. НС має чотири входи, на які подаються:

1. Обсяг інформації (кілобайт), що завантажується з бази даних за контрольний період. Отримане значення необхідно нормалізувати, оскільки зчитаний обсяг даних заздалегідь не відомий і індивідуальний для кожного завдання і для кожного користувача. Як нормалізацію можна застосувати оцінку трафіку за десятибальною шкалою (0 - обсяг дорівнює нулю, 10 - максимальний обсяг трафіку).

2. Кількість транзакцій в хвилину.

3. Кількість операцій модифікації даних в хвилину. У цьому прикладі АРМ використовує «короткі транзакції», тобто в рамках однієї транзакції зазвичай буває 1-2 операції модифікації даних.

4. Ознаки звернення до словника бази даних. Більшість клієнтських АРМ до словника не звертаються, що відрізняє їх від засобів розробки та адміністрування. Ознаки будуть дискретними (0 - немає звернень, 1 - є), і їх буде декілька - по одному на кожну з таблиць словника бази.

У даному прикладі використовується двошарова нейронна мережа прямого поширення, що містить один прихований шар з двох нейронів і вихідний шар з одним нейроном.

Навчання НС можна зробити за допомогою існуючих пакетів (наприклад, пакет Deductor Lite; MATLAB Neural Network Toolbox) або відомих алгоритмів (наприклад, метод «зворотного поширення помилки»). Для якісного навчання такої мережі необхідно близько 300 навчальних прикладів [5]. Слід зазначити, що підготовка навчальної вибірки є досить складним етапом.

Вихід НС може бути інтерпретований як процентна відповідність поточних дій діям хакера. Таким же чином можна організувати визначення різних атак та адаптацію до нових типів загроз.

Іншим прикладом використання НС в системах мережевої безпеки є нейроаналізатор, що входить до складу антивірусної утиліти AVZ. Нейроаналізатор дозволяє досліджувати підозрілі файли і застосовується в детекторі клавіатурних шпигунів (Keylogger).

Використання нейромережевих технологій дозволяє надати системам безпеки здатність до навчання, забезпечує високу точність розпізнавання. Недоліком є складність аналізу, внаслідок чого навчена НС представляється користувачеві «чорним ящиком» з певною кількістю входів і виходів.

На відміну від продукційних систем, зберігання нейронної мережі в обчислювальній машині вимагає набагато менше пам'яті, а визначення шкідливих дій - менше обчислювальних ресурсів. Ефект від цих переваг посилюється, якщо врахувати, що розробники прагнуть мінімізувати розмір оновлень для своїх систем безпеки.

Метод опорних векторів. Метод опорних векторів (Support vector machines, SVM) був описаний в роботах В. Н. Вапник [6, 7]. SVM - це математичний метод отримання функції, що вирішує завдання класифікації.

Ідея методу виникла з геометричної інтерпретації задачі класифікації. Нехай дві безлічі точок можна розділити площиною (в двовимірному просторі - прямою). Тоді таких площин буде нескінченна безліч (рис. 3а).

Виберемо в якості оптимальної таку площину, відстані до якої найближчих точок обох класів рівні (рис. 3б). Найближчі точки-вектори називаються *опорними*.

Пошук оптимальної площині приводить до задачі квадратичного програмування при безлічі лінійних обмежень-нерівностей. У 90-х рр.. минулого століття метод SVM був удосконалений: розроблені ефективні алгоритми пошуку оптимальної площині, знайдені способи узагальнення на нелінійні випадки і ситуації з числом класів, більшим двох [6, 7].

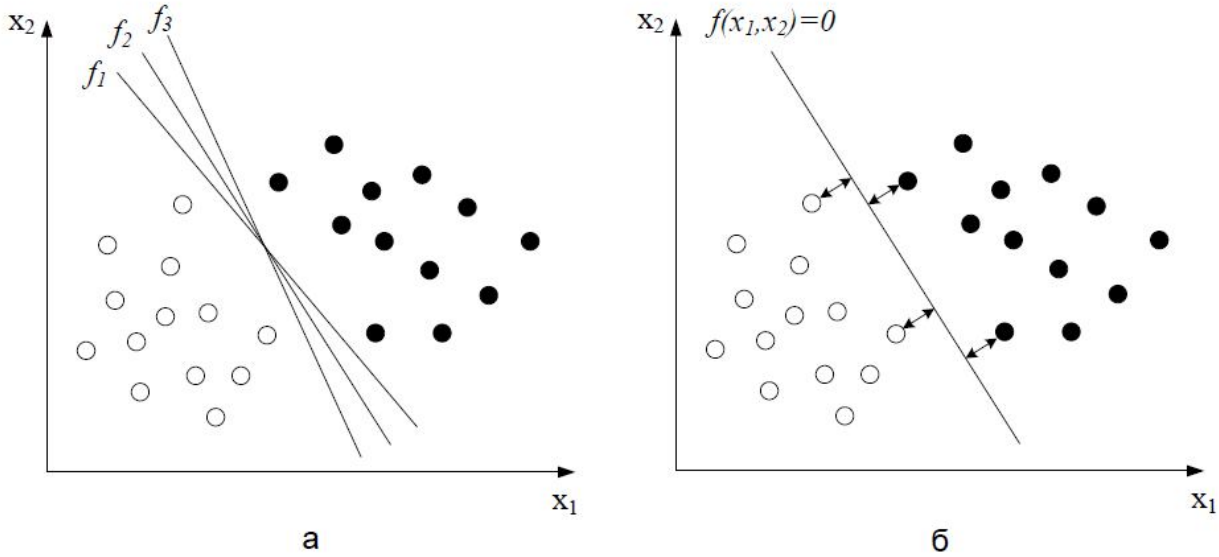

Рис. 2 Ілюстрація основної ідеї SVM

Метод опорних векторів добре зарекомендував себе в розпізнаванні рукописного тексту та облич, в задачах текстової класифікації. Ведуться розробки з використання цього методу в системах забезпечення інформаційної безпеки. Наприклад, в [8] описана методика визначення небажаного програмного забезпечення по метричної віддаленості від геометричного центру векторів-ознак подій комп'ютерної мережі за допомогою методу опорних векторів.

Для визначення атак потрібно сформувати вектор ознак, подібний вектору, який формується для штучної нейронної мережі. Потім за допомогою спеціального програмного забезпечення, наприклад SVM Light, провести навчання SVM-класифікатора. В результаті

вийде функція, яка буде виробляти класифікацію векторів-ознак, тобто розпізнавати, до якого класу належить поточне дію ПЗ або користувача - правомірному або забороненого.

Методи використання та навчання SVM у сфері мережевої безпеки ще до кінця не вивчені. Ясно тільки, що даний підхід має суттєві потужності і має великі перспективи розвитку, в тому числі в задачі забезпечення захисту комп'ютерних мереж.

Висновки. У нових антивірусних утилітах, програмах аналізу мережевої захищеності, міжмережевих екранах спостерігається тенденція збільшення масштабу використання технологій штучного інтелекту. Цьому сприяє наявність в них можливості навчання, активний розвиток методології ШІ, збільшення числа і ускладнення мережевих загроз.

Іншою тенденцією є спрямованість на інтеграцію засобів захисту різних рівнів (наприаєсклад, персональний антивірус і брандмауер рівня підприємства) з використанням засобів ШІ. Таким чином, можна зробити висновок, що розглянуті в статті підходи та методи ШІ на сьогоднішній день далеко не вичерпали свій потенціал. Висока ймовірність, що подальші дослідження розкриють нові шляхи застосування методів ШІ в сфері антивірусної безпеки.

ЛИТЕРАТУРА:

1. Люгер Д. Ф. Искусственный интеллект, стратегии и методы решения сложных проблем – Вильямс, 2003. – 864 с.
2. Рассел С., Норвиг П. Искусственный интеллект: современный подход. М.: Вильямс, 2007. – 1408 с.
3. Friedman N., Geiger D., Goldszmidt M., Bayesian Network Classifiers // Machine Learning. 1997. 29. P. 131–165.
4. Хайкин С. Нейронные сети: полный курс. 2-е изд. М.: Издательский дом Вильямс, 2006. – 1104 с.
5. Зайцев О. Нейросети в системах безопасности [Текст] // IT-Спец. – 2007. – № 6. – С. 54–59.
6. Vapnik V. Statistical learning theory. Wiley, New York, 1998.
7. Vapnik V. N. The Nature of Statistical Learning Theory. Springer-Verlag, 1995.

Lashkov P., Schäfer C., Kotenko I. Intrusion Detection in Unlabeled Data with Quarter-Sphere Support Vector Machine

Надійшла: 11.05.2012

Рецензент: д.т.н., проф. Юдін О.К.