РОЗРОБКА МЕТОДУ ПОБУДОВИ ДЕРЕВА З БАГАТОВИМІРНИМИ ВУЗЛАМИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ЗАДАЧ ІДЕНТИФІКАЦІЇ СТАНУ КОМП'ЮТЕРНОЇ СИСТЕМИ

аспірант В.В. Челак, д-р техн. наук, проф. С.Ю. Гавриленко,
Національний технічний університет "Харківський політехнічний інститут", м. Харків

Кожного дня зростають кількість інформації та її цінність. Слідуючи з цього зростає й кількість зловмисників, які розробляють нові загрози з метою модифікації, видалення, шифрування та виконання інших шкідливих дій з даними та комп'ютерною системою. Таким чином, підвищуються й вимоги до якості та швидкодії методів ідентифікації стану комп'ютерної системи з метою запобігання та виявлення загроз, вразливостей, шкідливого програмного забезпечення. Однак, не завжди існуючі методи здатні в повній мірі відповідати цим вимогам.

У роботі запропоновано новий метод побудови дерева рішень, який поєднує класичну модель побудови дерева рішень та оснований на щільності метод просторової кластеризації для вихідних даних з шумами (DBSCAN).

Основна ідея такого розбиття полягає в тому, що для критеріїв, які є об'єктами, тобто мають одразу декілька характеристик проводиться кластеризація методом DBSCAN. Надалі, обирається кластер з найбільшою кількістю об'єктів та розраховується центр цього кластеру θ та радіус гіперсфери ε, який визначає межі кластеру та надалі використовуються у якості . порогового значенням вузла рішень.

Процес побудови таких дерев рішень полягає в послідовному, рекурсивному розбитті навчальної множини на підмножини з застосуванням вирішальних правил в вузлах або за рахунок порівняння з пороговими значеннями на одновимірному просторі або за рахунок визначення приналежності до гіперсфери меншого порядку.

Результати моделювання показали, що запропонований метод надає можливість зменшити кількість розгалужень в дереві рішень, що дозволяє підвищити оперативність ідентифікації стану комп'ютерної системи. Використання належності до гіперсфер у якості критерію прийняття рішень надає можливість підвищити точність ідентифікації за рахунок нелінійності площині розбиття. Крім того, наявність більшої кількості гіперпараметрів надає можливість виконати більш оптимальне налаштування класифікатору. Такий метод є особливо ефективним за наявності вихідних даних, які мають високі кореляційні коефіцієнти так як об'єднує їх в один або декілька багатомірних критеріїв. Недоліком даного методу є збільшення часу навчання класифікатора. Крім того, такий метод потребує більшого об'єму пам'яті.