

## РОЗРОБКА ЕФЕКТИВНОГО КЛАСИФІКАТОРА ДАНИХ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ УПРАВЛІННЯ

Вінницький національний технічний університет;

### Анотація

В роботі розроблено ефективний метод класифікації даних в інтелектуальних системах управління. Процедури навчання класифікатора SVM покращена шляхом застосування пошуку опорних векторів за принципом поверхневого натягу замість стандартної оптимізаційної процедури.

**Ключові слова:** класифікація, машина опорних векторів (SVM), принцип поверхневого натягу, оптимізаційна процедура, помилка класифікації.

### Abstract

An effective method of data classification in intelligent control systems is developed in the work. The training procedures for the SVM classifier were improved by applying surface tension principle vector search instead of the standard optimization procedure.

**Keywords:** classification, support vector machine (SVM), surface tension principle, optimization procedure, classification error.

### Вступ

Однією з невирішених до кінця задач оптимальної стратегії розпізнавання [1] образів в інтелектуальних системах управління є побудови таких дискримінантних функцій в просторі ознак, які б давали максимально можливу точність класифікації за заданих умов розпізнавання, зокрема, в умовах нелінійного розміщення границь кластерів (зображень об'єктів) в просторі ознак.

Одним із найкращих класифікаторів, здатних реалізувати як лінійну, так і нелінійну класифікацію, визнано машину опорних векторів (SVM) [2]. Метод опорних векторів зводить навчання класифікатора до оптимізаційної задачі, яка розв'язується евристичними алгоритмами. Недоліком такого методу є складність обчислення опорних точок (кінців векторів), через які проводяться граничні для класів гіперплощини, між якими будується розподільна (дискримінантна) гіперплощина. Для пошуку опорних точок необхідно розв'язувати оптимізаційну задачу у вигляді лагранжіана [3].

Необхідність багатократного обчислення скалярного добутку векторів і пошуку оптимального значення функції Лагранжа породжує обчислювальну складність SVM методу. Тому метою даної роботи автори ставлять усунення обчислювальної складності SVM шляхом використання аналогії з фізичним законом утворення поверхневого натягу рідин для побудови класифікатора.

### Результати дослідження

"Машина (метод) опорних векторів" (SVM) - це контрольований алгоритм навчання, який використовується в основному для задач класифікації. У цьому алгоритмі кожен елемент даних подається як точка в  $n$ -мірному просторі (де  $n$  - це кількість ознак), причому величина кожної ознаки є значенням певної координати. SVM виконує класифікацію, знайшовши гіперплощину, яка розділяє два класи найкращим чином. Опорними векторами називаються об'єкти вибірки, що лежать на межах класів. Найкращий (оптимальний) розподіл між класами буде у випадку максимальної відстані між граничними гіперплощинами, що лежать на опорних точках. В більшості випадків границі між класами даних не є лінійними, однак формування узагальненої дискримінантної гіперплощини шляхом вибору нелінійної функції ядра (kernel function) дозволяє звести такі випадки до лінійних. Тому в даній роботі всі обґрунтування орієнтовані на лінійний класифікатор.

Загальний вигляд лінійної гіперплощини задається формулою

$$d(\vec{x}) = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n = w_0 + \vec{w}_a^T \vec{x},$$

де вектор  $\vec{w}_a = (w_1, w_2, \dots, w_n)^T$  називається ваговим або параметричним, а вектор  $\vec{x} = (x_1, x_2, \dots, x_n)^T$  - вектор ознак об'єкта.

Задачу усунення обчислювальної складності SVM автори вирішують шляхом використання аналогії з фізичним законом утворення поверхневого натягу рідин для побудови класифікатора. Знаходяться поверхневі точки кожного кластера, в якості яких вибираються неврівноважені силами взаємного притягання оточуючими точками (подібно поверхневим молекулам рідини), з цих поверхневих точок вибирають точки класів з найменшими відстанями, які і будуть опорними. Евристичний алгоритм пошуку поверхневих точок кластера складається з наступних кроків:

1 - Знаходяться відстані між всіма парами точок  $d(\vec{x}_i, \vec{x}_j)$ ;

2 - За цими відстанями для кожної точки вибираємо оптимальну кількість  $t$  точок, найближчих до вибраної; експериментально встановлено  $t = 4..6$ ;

3 - Центр координат розміщується в заданій точці  $\vec{x}_k$  шляхом паралельного перенесення координат вибраної підмножини точок;

4 - Підраховується залишкова сила  $f(\vec{x}_p) = \frac{\{(x_{ik} + x_{jk})\}}{d(\vec{x}_i, \vec{x}_j)}, k = 1, \dots, n$  поверхневого натягу

вибраної точки  $\vec{x}_p$  шляхом знаходження рівнодійної сил притягання від усіх точок вибраної підмножини. При цьому використовуються тільки операції алгебраїчного додавання пронормованих координат точок. Якщо  $f_p \geq f_{nop}$ , то точка вважається поверхневою.

5 - Поверхневі точки двох кластерів з найменшими відстанями вважаються опорними, і через них проводяться опорні гіперплощини, між якими розміщується оптимальна з точки зору мінімізації помилки класифікації гіперплощина.

Як показали експериментальні дослідження розробленого алгоритму, різниця між знайденими координатами опорних точок запропонованим методом і класичним методом складає у відносному плані 3%, що в незначній мірі впливає на точність роботи класифікатора SVM. При цьому швидкість класифікації даних збільшується на порядок.

## Висновки

В роботі запропоновано метод підвищення ефективності класифікації даних в інтелектуальних системах управління шляхом вдосконалення процедури навчання класифікатора SVM. В результаті застосування пошуку опорних векторів за принципом поверхневого натягу замість стандартної оптимізаційної процедури швидкість роботи класифікатора збільшилась на порядок при збереженні притаманної класифікатору точності класифікації. Проведені експериментальні дослідження розроблених алгоритмів класифікації підтвердили адекватність запропонованого методу.

## Список використаної літератури

1. Bykov N.M., Kuzmin I.V., Yakovenko A.I. Development of effective strategy of pattern recognition // Proceedings of SPIE. - 2000, Vol. 4425, pp. 76-82.
2. Константин Воронцов. Лекция по методу опорных векторов <http://www.ccas.ru/voron/download/SVM.pdf>
3. Маннинг К.Д., Рагхаван П., Шютце Х. Введение в информационный поиск.— М.: Изд-во “Вильямс”, 2011.

**Богдан Олегович Волоський** — студент групи КІВ-166, факультет комп'ютерних систем та автоматики, Вінницький національний технічний університет, м.Вінниця, e-mail: [bogdan.volosky@gmail.com](mailto:bogdan.volosky@gmail.com).

**Микола Максимович Биков** — професор кафедри комп'ютерних систем управління, Вінницький національний технічний університет, м. Вінниця, e-mail: [nkbykov@vntu.edu.ua](mailto:nkbykov@vntu.edu.ua).

**Bogdan J. Voloskyi** — student of Computer System and Automation Department, KIV-16b group, Vinnytsia National Technical University, Vinnytsia, e-mail: [bogdan.volosky@gmail.com](mailto:bogdan.volosky@gmail.com).

**Mykola M. Bykov** — professor of Computer Control System Department, Vinnytsia National Technical University, Vinnytsia, e-mail: [nkbykov@vntu.edu.ua](mailto:nkbykov@vntu.edu.ua).