

ЗАСТОСУВАННЯ МЕТОДУ ОПОРНИХ ВЕКТОРІВ ДЛЯ ВИЗНАЧЕННЯ НЕОБХІДНОСТІ ШТУЧНОГО ПОЛИВУ РОСЛИН НА ОСНОВІ ІНДУКЦІЇ ФЛУОРЕСЦЕНЦІЇ ХЛОРОФІЛУ

Вступ. Індукція флуоресценції хлорофілу (ІФХ) – це спосіб отримання додаткової інформації про стан рослин без їх пошкодження. ІФХ виникає при освітленні рослини у синьому спектрі світла. При цьому отримується так звана крива Каутського у червоному спектрі світла [1, 2].

В даний час існують роботи з визначення водного дефіциту на основі ІФХ з використанням нейронних мереж [3–5]. В даній роботі пропонується оцінити можливість визначення необхідності штучного поливу рослин сої на основі методу опорних векторів (Support Vector Machine, SVM) і порівняти ці результати з результатами, отриманими автором при аналізі ІФХ цинії у [6].

Метод опорних векторів здатний виконувати лінійну або нелінійну класифікацію, регресію і навіть знаходження викидів у даних. Основна ідея методу – перетворення вхідних векторів у простір більш високої розмірності і пошук розділяючої гіперплощини з максимальним проміжком у цьому просторі. Перетворення вхідних векторів відбувається за допомогою спеціальних функцій – ядер. Дві паралельні гіперплощини будуються за двома сторонами гіперплощини, що розділяє класи. Розділяючою гіперплощиною буде гіперплощина, що максимізує відстань до двох паралельних гіперплощин. Алгоритм працює за припущенням, що чим більша різниця або відстань між цими паралельними гіперплощинами, тим менша середньоквадратична похибка класифікації.

Найбільш поширені ядра – лінійне ядро, поліноміальне ядро, гауссове RBF ядро, сигмоїдне ядро [7]. Підбір оптимального ядра для наявного набору даних може покращити результат класифікації або регресії.

В мові Python реалізації методу опорних векторів надаються бібліотекою Scikit-learn класами SVC [8] (для класифікації) та SVR [9] (для регресії).

Досліджено можливості методу опорних векторів (SVM) при визначенні необхідності штучного поливу на основі індукції флуоресценції хлорофілу. Визначено найкраще ядро SVM для згаданої цілі та найкращий метод зниження розмірності виміряних даних індукції флуоресценції хлорофілу.

Ключові слова: метод опорних векторів, індукція флуоресценції хлорофілу, зниження розмірності даних.

1. Матеріали та методи. Для експерименту використано рослини сої, які було висаджено у горщики і поділено на три групи. Перша група отримувала полив 50 мл/кг ґрунту, друга 150 мл/кг ґрунту, третя група була без поливу. Перед початком вимірювання ІФХ, всі рослини були однаково рясно политі. Вимірювання ІФХ відбулось перед початком експерименту і далі впродовж трьох тижнів у лабораторних умовах (всього 12 серій вимірювання). В кінці експерименту рослини сої були у нормальному стані, не зів'ялі. При цьому температура у лабораторії підтримувалася в межах 19–24°C. За годину до вимірювання рослини поміщалися у тінь, що дозволило зменшити темнову адаптацію рослин перед вимірюванням ІФХ з 15 хв, до 4 хв.

Вимірювання ІФХ здійснювалося чотирма сенсорами ІФХ, розробленими в Інституті кібернетики імені В.М. Глушкова НАН України, що вимірювали зміну ІФХ впродовж 5 хв. Вимірювання температури, кислотності та вологості ґрунту проводилось з допомогою аналізатора ґрунту КС-300. Категоріальні значення вологості ґрунту, що видавав прилад, переводилися у цифрові, шляхом заміни категоріальних значень відповідними цифровими позначеннями (від 1 до 5). Кожен параметр у кожному горщику з рослинами вимірювався три рази у різних місцях. Отримані значення усереднювались. Також додатково замірялась температура повітря при кожному вимірюванні кривої ІФХ.

Для аналізу даних використано реалізацію SVM для регресії у Python, а саме клас SVR (Support Vector Regression) з метою прийняття рішення про штучний полив. Крім значень флуоресценції хлорофілу (F) на вхід SVR пропонується подавати температуру повітря (T_{air}), оскільки відомо, що деякі ділянки кривих залежать від неї. Модель SVR за вхідними даними має встановити вологість ґрунту (H_{gr}). Таким чином SVR здійснює регресію такого вигляду:

$$\varphi(F, T_{air}) = H_{gr}.$$

Для оцінки результату регресії використано такі критерії якості, як коефіцієнт детермінації (R^2) та корінь середньоквадратичної похибки ($RMSE$).

Коефіцієнт детермінації можна обчислити за формулою:

$$R^2 = 1 - \frac{\sigma^2}{\sigma_h^2},$$

де σ_h^2 – дисперсія виміряних значень вологості, σ^2 – дисперсія похибки моделі.

В ідеалі коефіцієнт детермінації – це число переважно в інтервалі $[0;1]$, проте якщо якість моделі погана, то число коефіцієнт детермінації може виходити за межі цього інтервалу залежно від формули, яка застосовується для його обчислення. При цьому коефіцієнт детермінації дуже часто буває від'ємним. В даному дослідженні він обчислювався з допомогою функції Python `sklearn.metrics.r2_score`. Чим ближче коефіцієнт детермінації до 1, тим кращим вважається результат регресії. Взнявши корінь з коефіцієнту детермінації, можна отримати коефіцієнт кореляції R .

Корінь середньоквадратичної похибки отримується за формулою:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2},$$

де Y_i – початкові виміряні значення, \hat{Y}_i – визначені моделлю значення, n – кількість виміряних значень. Чим менший корінь середньоквадратичної похибки, тим кращим вважається результат. $RMSE$ виражається у тих же одиницях, що і виміряні значення.

На вхідних даних тестування SVR відбувалося 1000 разів. При цьому кожен раз відбувалось випадкове розбиття даних на навчальну і тестову вибірку. На тестову вибірку відводилось 10 % вимірювань із 432 вимірювань ІФХ. В результаті тестування зберігалась інформація про мінімальний коефіцієнт детермінації (R_{\min}^2), середній (R_{avg}^2), максимальний (R_{\max}^2) і відповідні корені середньоквадратичних похибок ($RMSE_{\min}$, $RMSE_{avg}$, $RMSE_{\max}$).

2. Використання різних ядер SVM. Спочатку оцінювалися можливості SVR при використанні трьох ядер (поліноміального, гауссового ядра з радіально-базовою функцією та сігмоїдного). Лінійне ядро не використовувалося через незадовільні результати і значний час навчання алгоритму на базі даного ядра. Дані ІФХ брались без нормалізації. Поліноміальне ядро – третього степеня. Найкращий результат ($R_{\max}^2 = 0.28$) був отриманий при поліноміальному ядрі. Тестування моделі SVR при степенях поліноміального ядра 2, 3, 4, 5 показало, що найкращий результат був отриманий при використанні поліноміального ядра четвертого степеня ($R_{\max}^2 = 0.4096$), тому в подальшому тестування проводилось на основі даного ядра з цим степенем.

3. Нормалізація вхідних даних. Покращити результат роботи методу можна з допомогою нормалізації вимірювань ІФХ. Для цього можна скористатися мінімаксною нормалізацією у межах $[-1, 1]$ та z -score нормалізацією [6]. Як видно з табл. 1, мінімаксна нормалізація дає незначне покращення.

ТАБЛИЦЯ 1. Результат навчання SVR моделі за різними способами нормалізації

Вид нормалізації	Критерії якості					
	R_{\min}^2	R_{avg}^2	R_{\max}^2	$RMSE_{\min}$	$RMSE_{avg}$	$RMSE_{\max}$
Без нормалізації	- 0.77	- 0.06	0.40	0.73	1.05	1.42
z -score	- 0.62	- 0.05	0.19	0.69	1.05	1.35
<i>minmax</i>	- 0.80	0.02	0.41	0.64	1.01	1.32

4. Зниження розмірності даних. Навчання моделі SVR також було вивчено при взятті нерівномірно різної кількості точок на кривій ІФХ за ступеневою шкалою (ступінь 1/8). Початково дані, виміряні сенсорами ІФХ, містили 90 точок. Використовувалась мінімаксна нормалізація вимірювань ІФХ та поліноміальне ядро четвертого степеня. Результат регресії при взятті різної кількості точок наведено в табл. 2.

ТАБЛИЦЯ 2. Результат навчання SVR моделі із взяттям різної кількості точок на кривій ІФХ

Кількість точок	Критерії якості					
	R_{\min}^2	R_{avg}^2	R_{\max}^2	$RMSE_{\min}$	$RMSE_{avg}$	$RMSE_{\max}$
10	- 0.76	0.04	0.43	0.67	0.99	1.29
20	- 0.50	0.03	0.36	0.68	1.01	1.34
30	- 0.68	0.03	0.37	0.62	1.01	1.37
40	- 0.54	0.03	0.38	0.72	1.00	1.32
50	- 0.54	0.03	0.38	0.73	1.00	1.27
60	- 0.71	0.03	0.39	0.68	0.99	1.37
70	- 0.48	0.03	0.40	0.71	1.00	1.29
80	- 0.65	0.04	0.40	0.66	0.99	1.31
90	- 0.68	0.03	0.41	0.71	1.00	1.34

Як видно з табл. 2, найкращий результат отриманий із взяттям 10 точок на кривій ІФХ за степенною шкалою (ступінь 1/8). Отриманий найкращий результат регресії показано на рис. 1.

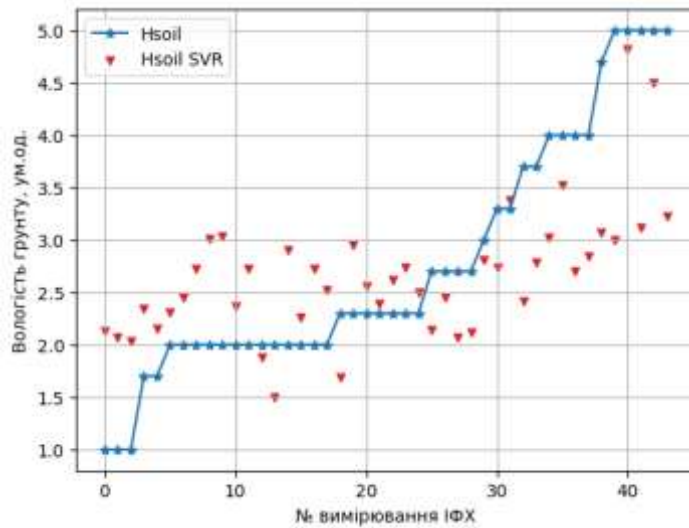


РИС. 1. Найкращий результат регресії із взяттям 10 точок на кривій ІФХ

В наукових публікаціях [3–5] також можна зустріти й інші методи зниження розмірності даних. Зокрема використання різних параметрів кривої ІФХ таких, як F_0 , F_s , F_m і т. п. Також можна застосувати метод головних компонент PCA та використання як вхідний вектор коефіцієнтів апроксимуючого полінома виду[10]:

$$Q(x) = c_0x^n + c_1x^{n-1} + \dots + c_{n-1}x + c_n,$$

де c_0 – завжди рівний 1, як x бралися порядкові значення відліків, $x \in Z$, $x = 1 \dots 90$, оскільки із взяттям за x відліків реального часу апроксимація проходить неякісно. Результат застосування різних способів зниження розмірності наведено в табл. 3.

ТАБЛИЦЯ 3. Результати SVR моделі при зниженні розмірності різними способами

Вхідний вектор	Критерії якості					
	R_{\min}^2	R_{avg}^2	R_{\max}^2	$RMSE_{\min}$	$RMSE_{\text{avg}}$	$RMSE_{\max}$
10 точок за степенною шкалою, T_{air}	-0.76	0.04	0.43	0.67	0.99	1.29
$F_m, F_s, F_0, F_j, F_i, F_v, Area,$ $F_m / F_s, F_m / F_v, (F_m - F_s) / F_s,$ $(F_m - F_s) / F_m, Rfd, V_j, V_i, V_s, T_{\text{air}}$	-7.83	-0.69	0.16	0.83	1.29	3.15
$F_0, F_i, F_j, F_m, F_s, T_{\text{air}}$	-2.99	-0.36	0.20	0.66	1.19	1.88
6 головних компонент, виданих методом головних компонент, T_{air}	-7.54	-0.37	0.18	0.79	1.19	2.75
$c_0 \dots c_9, T_{\text{air}}$	-2.98	-0.26	0.2313	0.69	1.14	2.10

З табл. 3 ми бачимо, що краще використовувати базові параметри ІФХ (F_0 , F_i , F_j , F_m , F_s) ніж великий набір похідних параметрів. При використанні методу головних компонент бралась різна кількість отриманих головних компонент. Найкраще регресія отримувалась із взяттям 6 головних компонент. З поміж усіх методів зниження розмірності найкраще себе демонструє метод взяття 10 точок на кривій за степеневою шкалою (ступінь 1/8) з мінімаксною нормалізацією вхідного вектора моделі SVR. Другий за ефективністю метод полягає у використанні коефіцієнтів апроксимуючого полінома.

В роботі [6] при використанні метод опорних векторів на даних поливу цинії, при відведенні 20 % вимірювань ІФХ на тестову вибірку було отримано $R_{\max}^2 = 0,54$. В даному дослідженні при відведенні 20 % вимірювань ІФХ на тестову вибірку було отримано $R_{\max}^2 = 0,35$. Гірший результат можна пояснити тим, що соя є посухостійкою рослиною, а тому дефіцит вологи менше впливає на зміни ІФХ.

Висновки. Дослідження методу опорних векторів з метою визначення необхідності штучного поливу на основі кривих ІФХ показало, що найкращий результат отримується при використанні поліноміального ядра четвертого степеня з використанням 10 точок кривої ІФХ, взятих нерівномірно за степеневою шкалою (ступінь 1/8) з використанням мінімаксної нормалізації вимірювань ІФХ. Також показано, що метод опорних векторів дає гірші результат при аналізі кривих ІФХ рослин сої ніж при аналізі вимірювань ІФХ цинії, що можна пояснити тим, що соя посухостійка рослина, а тому ІФХ у меншій мірі відображає дефіцит вологи у ґрунті.

Список літератури

1. Kalaji H.M., Schansker G., Ladle R.J. et al Frequently asked questions about in vivo chlorophyll fluorescence: practical issues. *Photosynthesis Research*. 2014. Vol. 122, Iss. 2. P. 121–158. <https://doi.org/10.1007/s11120-014-0024-6>
2. Kalaji H.M. et al. Frequently asked question about chlorophyll fluorescence, the sequel. *Photosynthesis Research*. 2017. Vol. 132, Iss. 1. P. 13 – 66. <https://doi.org/10.1007/s11120-016-0318-y>
3. Goltsev V., Zaharieva I., Chernev P., Kouzmanova M., Kalaji H.M., Yordanov I., Krasteva V., Alexandrov V., Stefanov D., Alkhverdiev S.I., Strasser R.J. Drought-induced modification of photosynthetic electron transport in intact leaves: Analysis and use of neural network as a tool for a rapid non-invasive estimation. *Biochimica et Biophysica Acta*. 2012. Vol. 1817, Iss. 8. P. 1490–1498. <https://doi.org/10.1016/j.bbabi.2012.04.018>
4. Zhou Ch., Mao J.D., Zhao H., Zhang B. Combining principle component analysis and a back-propagation network for classifying plant water stress based on images of chlorophyll fluorescence. *Frontier research and Innovation in Optoelectronics technology and Industry*. London: Taylor&Francis Group, 2019. P. 241–248.
5. Rybka K., Janaszek-Mankowska M., Siedlarz P., Mankowski D. Machine learning in determination of water saturation deficit in wheat leaves on basis of Chl a fluorescence parameters. *Photosynthetica*. 2019. Vol. 57, Iss. 1. P. 226–230. <http://dx.doi.org/10.32615/ps.2019.017>
6. Груша В. Інтелектуальна обробка даних від хлорофіл-флуорометричних сенсорів. *Кібернетика та комп'ютерні технології*. 2022. № 1. С. 42–48. <https://doi.org/10.34229/2707-451X.22.1.5>
7. Srivastava D.K., Bhambhu L. Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology*. 2010. Vol 12, No. 1. <https://www.jatit.org/volumes/Vol12No1/1Vol12No1.pdf>
8. Scikit-learn: sklearn.svm. SVC. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html> (звернення: 25.03.2024).
9. Scikit-learn: sklearn.svm. SVR. https://scikit-learn.org/stable/modules/_generated/sklearn.svm.SVR.html#sklearn.svm.SVR (звернення: 25.03.2024).
10. NumPy: numpy.poly. <https://numpy.org/doc/stable/reference/generated/numpy.poly.html> (звернення: 25.03.2024)

Одержано 09.04.2024

Груша Володимир Михайлович,

кандидат технічних наук, старший науковий співробітник
Інституту кібернетики імені В.М. Глушкова НАН України, Київ.

<https://orcid.org/0000-0002-2497-0939>

vhrusha@gmail.com

УДК 004.94

В.М. Груша

Застосування методу опорних векторів для визначення необхідності штучного поливу рослин на основі індукції флуоресценції хлорофілу

Інститут кібернетики імені В.М. Глушкова НАН України, Київ

Листування: vhrusha@gmail.com

Вступ. Індукція флуоресценції хлорофілу (ІФХ) – це спосіб отримання додаткової інформації про стан рослин без їх пошкодження. ІФХ виникає при освітленні рослини у синьому спектрі світла. При цьому отримується так звана крива Каутського у червоному спектрі світла. В даний час існують роботи з визначення водного дефіциту на основі ІФХ з використанням нейронних мереж. Метод опорних векторів (SVM) – модель машинного навчання, здатна виконувати лінійну або нелінійну класифікацію, регресію і навіть знаходження викидів у даних. SVM може виступати альтернативою до нейронних мереж при аналізі вимірювань індукції флуоресценції хлорофілу.

Мета роботи – оцінити можливість визначення необхідності штучного поливу рослин сої на основі методу опорних векторів і порівняти ці результати з результатами, отриманими автором в більш-ранніх експериментах при аналізі ІФХ рослин цинії.

Результати. Здійснено дослідження SVM з використанням різних ядер, різних способів нормалізації вимірювань ІФХ, різних способів зменшення розмірності даних. Використовувалася реалізація SVM з бібліотеки Scikit-learn, клас SVR (Support Vector Regression). Експериментально встановлено найкраще ядро, найкращий спосіб нормалізації та спосіб формування вхідного вектора для методу опорних векторів з метою визначення необхідності штучного поливу.

Висновки. Дослідження методу опорних векторів з метою визначення необхідності штучного поливу на основі кривих ІФХ показало, що найкращий результат отримується при використанні поліноміального ядра четвертого степеня з використанням 10 точок кривої ІФХ, взятих нерівномірно за степеневою шкалою (ступінь 1/8) з використанням мінімаксної нормалізації вимірювань ІФХ. Також показано, що SVM дає гірші результати при аналізі кривих ІФХ рослин сої ніж при аналізі вимірювань ІФХ цинії, що можна пояснити тим, що соя посухостійка рослина, а тому ІФХ у меншій мірі відображає дефіцит вологи у ґрунті.

Ключові слова: метод опорних векторів, індукція флуоресценції хлорофілу, зниження розмірності даних.

UDC 004.94

Volodymyr Hrusha

Using Support Vector Machine for Determining the Need for Artificial Watering Based on the Chlorophyll Fluorescence Induction

V.M. Glushkov Institute of Cybernetics of the NAS of Ukraine, Kyiv

Correspondence: vhrusha@gmail.com

Introduction. Chlorophyll fluorescence induction (CFI) is a method for receiving additional information about state of plants without its injuring. CFI occurs when the plant is illuminated in the blue spectrum of light. The so-called Kautsky curve in the red spectrum of light is measuring. Currently, there are scientific papers about determining water deficit using CFI and neural networks. Support Vector Machine (SVM) – the model of

machine learning capable of performing linear or non-linear classification, regression and even finding outliers in data. SVM can be an alternative to neural networks for analysis of CFI measurements.

The purpose of the paper is to assess the possibility of determining the need for artificial watering of soybean plants based on the SVM and compare the results with the results obtained by the author in earlier experiments of analyzing the CFI of zinnia plants.

Results. SVM research was conducted using different SVM kernels, different methods of normalizing CFI measurements, different methods of the dimensionality reduction of CFI data. The SVM implementation from the Scikit-learn Python library, the SVR (Support Vector Regression) class was used. The best kernel of SVM, the best normalization method, and the best method of forming the input vector for the support vector machine were experimentally revealed for determining the need of artificial watering.

Conclusions. The research of Support Vector Machine for the purpose of determining the need for artificial watering based on CFI curves were conducted. The research showed that the best result is obtained (1) using a polynomial kernel of the fourth degree, (2) using 10 points of CFI curve taken unevenly on an exponentiation scale (power of 1/8), (3) using minimax normalization of CFI measurements. SVM showed a worse result in the analysis of CFI curves of soybean plants than in the analysis of CFI curves of the zinnia plants. It can be explained by the fact that the soybean is a drought resistant plant and therefore CFI demonstrated worse a moisture deficit.

Keywords: Support Vector Machine, chlorophyll fluorescence induction, dimensionality reduction of data.