ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

УДК 519.71:519.712:631.52 https://doi.org/10.35546/kntu2078-4481.2021.2.9

О.В. ЛАРЧЕНКО
Херсонський державний аграрно-економічний університет
ORCID: 0000-0001-7857-0802
Г.О. ДИМОВА
Херсонський державний аграрно-економічний університет
ORCID: 0000-0002-5294-1756

ВИКОРИСТАННЯ МЕТОДУ ГОЛОВНИХ КОМПОНЕНТ І КАНОНІЧНИХ КОРЕЛЯЦІЙ ДЛЯ ВИВЕДЕННЯ НОВИХ СОРТІВ КУЛЬТУРНИХ РОСЛИН

Для виведення нових сортів культурних рослин у статті запропоновано використання методу головних компонент на першому етапі розрахунку та методу канонічних кореляцій на другому. Аналіз методів генетики та селекції з питання виведення нових сортів культурних рослин привів до висловлення наступних припущень: зміна основної властивості якогось виду рослини пов'язане з можливою величиною розкиду окремих часткових властивостей, а також найбільші зміни в бажаних властивостях можна отримати в результаті схрещування особин, що мають найбільшу варіабельність цих властивостей. Виходячи з цього при цілеспрямованому доборі слід враховувати в першу чергу лише ті ознаки, які виявляють найбільший розкид при переході від одного об'єкта до іншого, що можливо при використанні методу головних компонент.

Метод головних компонент є одним з основних способів зменшити розмірність даних, втративши найменшу кількість інформації. Проблема зводиться до класичної задачі знаходження власних значень та власних векторів симетричної матриці. Серед існуючих підходів розв'язання задачі методом головних компонент (факторів) в даній роботі використовується підхід, де для максимізування функції, яка зв'язана з деяким числом додаткових умов, користуємося методом множників Лагранжа.

Метод канонічних кореляцій на другому етапі дозволяє знаходити максимальні кореляційні зв'язки між двома групами випадкових величин, будучи як би узагальненням кореляційного аналізу на випадок двох множин випадкових величин. Ця залежність визначається за допомогою нових аргументів - канонічних величин (канонічних змінних), обчислених як лінійні комбінації вихідних ознак по кожній з груп. Канонічні величини повинні максимально корелювати між собою, а їх число визначається за кількістю змінних в меншій множині (якщо число змінних в них не однаково). Тобто на другому етапі визначається перебування канонічних кореляцій для знайдених пар головних компонент і вторинних векторів, елементами яких є знайдені головні компоненти. Таким чином розроблений алгоритм спрямованого пошуку пар для схрещування культурних рослин, що складається з двох етапів розрахунку методами головних компонент та канонічних кореляцій.

Ключові слова: селекція, генетика, спрямований пошук, головні компоненти, коваріаційна матриця, дисперсія.

О.В. ЛАРЧЕНКО
Херсонский государственный аграрно-экономический университет
ORCID: 0000-0001-7857-0802
А.О. ДЫМОВА
Херсонский государственный аграрно-экономический университет
ORCID: 0000-0002-5294-1756

ИСПОЛЬЗОВАНИЕ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ И КАНОНИЧЕСКИХ КОРРЕЛЯЦИЙ ДЛЯ ВЫВЕДЕНИЯ НОВЫХ СОРТОВ КУЛЬТУРНЫХ РАСТЕНИЙ

Для выведения новых сортов культурных растений в статье предложено использование метода главных компонент на первом этапе расчета и метода канонических корреляций на втором. Анализ методов генетики и селекции по вопросу выведения новых сортов культурных растений привел к высказыванию следующих предположений: изменение основного свойства какого-либо вида растения связано с возможной величиной разброса отдельных частных свойств, а также наибольшие изменения в желаемых свойствах можно получить в результате скрещивания особей, имеют наибольшую вариабельность этих свойств. Исходя из этого при целенаправленном подборе следует учитывать в первую очередь только те признаки, которые проявляют наибольший разброс при переходе от одного объекта к другому, что возможно при использовании метода главных компонент.

Метод главных компонент является одним из основных способов уменьшить размерность данных, потеряв наименьшее количество информации. Проблема сводится к классической задачи нахождения собственных значений и собственных векторов симметричной матрицы. Среди существующих подходов решения задачи методом главных компонент (факторов) в данной работе используется подход, где для максимизации функции, связанной с некоторым числом дополнительных условий, пользуемся методом множителей Лагранжа.

Метод канонических корреляций на втором этапе позволяет находить максимальные корреляционные связи между двумя группами случайных величин, являясь как бы обобщением корреляционного анализа в случае двух множеств случайных величин. Эта зависимость определяется с помощью новых аргументов - канонических величин (канонических переменных), вычисленных как линейные комбинации исходных признаков по каждой из групп. Канонические величины должны максимально коррелировать между собой, а их число определяется по количеству переменных в меньшем множестве (если число переменных в них не одинаково). То есть на втором этапе определяется пребывание канонических корреляций для найденных пар главных компонент и вторичных векторов, элементами которых являются найденные главные компоненты. Таким образом разработан алгоритм направленного поиска пар для скрещивания культурных растений, состоящий из двух этапов расчета методами главных компонент и канонических корреляций.

Ключевые слова: селекция, генетика, направленный поиск, главные компоненты, ковариационная матрица, дисперсия.

O. LARCHENKO
Kherson State Agrarian and Economic University
ORCID: 0000-0001-7857-0802
H. DYMOVA
Kherson State Agrarian and Economic University
ORCID: 0000-0002-5294-1756

USING THE MAIN COMPONENTS METHOD AND CANONICAL CORRELATIONS FOR BREEDING NEW VARIETIES OF CULTURAL PLANTS

To develop new varieties of cultivated plants, the article proposes the use of the method of principal components at the first stage of the calculation and the method of canonical correlations at the second. Analysis of the methods of genetics and breeding on the issue of breeding new varieties of cultivated plants led to the following assumptions: a change in the basic property of any plant species is associated with the possible magnitude of the scatter of individual particular properties, and the greatest changes in the desired properties can be obtained as a result of crossing individuals that have the greatest variability of these properties. Based on this, in purposeful selection, first of all, only those features should be taken into account that show the greatest scatter in the transition from one object to another, which is possible when using the method of principal components.

Principal component method is one of the main ways to reduce the dimension of data, losing the least amount of information. The problem is reduced to the classical problem of finding the eigenvalues and eigenvectors of a symmetric matrix. Among the existing approaches to solving the problem by the principal components (factors) method, this study used an approach where, to maximize a function associated with a number of additional conditions, we use the method of Lagrange multipliers.

The canonical correlations method at the second stage makes it possible to find the maximum correlations between two groups of random variables, being, as it were, a generalization of the correlation analysis in the case of two sets of random variables. This dependence is determined by using new arguments - canonical values (canonical variables), calculated as linear combinations of the original features for each of the groups. Canonical values should be maximally correlated with each other, and their number is determined by the number of variables in a smaller set (if the number of variables in them is not the same). That is, at the second stage, the residence of the canonical correlations is determined for the found pairs of principal components and secondary vectors, the elements of which are the found principal components. Thus, an algorithm for directed search for pairs for crossing cultivated plants has been developed, which consists of two stages of calculation by the methods of principal components and canonical correlations.

Keywords: selection, genetics, directed search, principal components, covariance matrix, variance.

Постановка проблеми

На основі аналізу методів генетики та селекції з питання виведення нових сортів культурних рослин можна зробити наступні припущення:

1) Зміна основної властивості якогось виду рослини пов'язане з можливою величиною розкиду окремих часткових властивостей;

2) Найбільші зміни в бажаних властивостях можна отримати в результаті схрещування особин, що мають найбільшу варіабельність цих властивостей.

Тому при цілеспрямованому доборі слід враховувати в першу чергу лише ті ознаки, які виявляють найбільшу мінливість (найбільший розкид) при переході від одного об'єкта до іншого.

Аналіз останніх досліджень і публікацій

Селекція – наука, що розробляє шляхи створення нових і поліпшення існуючих сортів рослин, порід тварин і штамів мікроорганізмів [1].

Створення нових сортів і порід ґрунтується на таких найважливіших властивостях живого організму, як спадковість і мінливість [2]. Саме тому генетика – наука про мінливість і спадковість організмів - є теоретичною основою селекції.

Маючи свої власні задачі і методи, селекція твердо спирається на закони генетики, є важливою областю практичного використання закономірностей, встановлених генетикою. Разом з тим селекція спирається і на досягнення інших наук. Різним напрямам селекції присвячені роботи таких вчених як І.В. Мічурін, Г.Д. Карпеченко, Б.Л. Астауров, Н.І. Вавілов. На сьогоднішній день генетика вийшла на рівень цілеспрямованого конструювання організмів з потрібними ознаками і властивостями.

Формулювання мети дослідження

Метою роботи є створення алгоритму спрямованого пошуку пар для схрещування культурних рослин з використанням методів головних компонент на першому етапі та знаходження канонічних кореляцій для знайдених пар головних компонент на другому етапі.

Викладення основного матеріалу дослідження

В практичній і дослідницькій роботі селекціонера приходиться стикатися з ситуаціями, коли загальне число P ознак $X^{(1)}$, $X^{(2)}$, …, $X^{(P)}$, що реєструється на кожному з множини досліджуваних об’єктів (рослин даного типу), дуже велико [3], тобто є багатовимірні спостереження

$$X_i = \begin{pmatrix} X_i^{(1)} \\ X_i^{(2)} \\ \vdots \\ X_i^{(P)} \end{pmatrix}, \tag{1}$$

які необхідно осмислити і вилучити потрібну інформацію.

Перше природне бажання будь-якого дослідника – це уявити багатовимірне спостереження у вигляді одного числа або хоча б у вигляді вектора **Z** деяких допоміжних показників $Z^{(1)}, Z^{(2)}, ..., Z^{(P')}$ з істотно меншим (ніж P) числом компонент P', обумовлено наступними причинами:

- необхідністю наочного уявлення вихідних даних, що досягається їх проектуванням на спеціально підібраний тривимірний простір ($P' = 3$), площину ($P' = 2$) або числову пряму;
- прагненням до лаконізму досліджуваних моделей;
- необхідністю істотного стиснення обсягів збереженої статистичної інформації без істотних втрат в її інформативності.

Є три основних типи принципових передумов, що обумовлюють можливість переходу від більшого числа P вихідних показників стану аналізованого об'єкта (рослини) до істотно меншого числа P найбільш інформативних змінних. Це, по-перше, дублювання інформації, що доставляється сильно взаємопов'язаними ознаками; по-друге, неінформативність ознак, мало мінливих при переході від одного об'єкта до іншого (мала варіабельність ознак); по-третє, можливість агрегування, тобто простого або «зваженого» підсумовування за деякими ознаками [3]. Наведемо формально завдання переходу (з найменшими втратами в інформативності) до нового набору ознак $Z^{(1)}, Z^{(2)}, ..., Z^{(P')}$. Нехай **Z** = $Z(X)$ – деяка P-мірна вектор-функція вихідних змінних $X^{(1)}$, $X^{(2)}$, …, $X^{(P)}$ ($P' << P$) і нехай $I_{P'}(Z(X))$ – певним чином задана міра інформативності P-мірної системи ознак $Z(X) = (Z^{(1)}(X), Z^{(2)}(X), \ldots Z^{(P')}(X))$. Конкретний вибір функціоналу $I_{P'}(Z)$ залежить від специфіки розв'язуваної задачі і спирається на один з можливих критеріїв:

- критерій автоінформатівності, націлений на максимальне збереження інформації, яка міститься у вихідному масиві $\{X_i\}, i = \overline{1, n}$ стосовно самих початкових ознак;
- критерій зовнішньої інформативності, націлений на максимальне «витиснення» з $\{X_i\}$ інформації, що міститься в цьому масиві відносно деяких інших (зовнішніх) показників.

Задача полягає у визначенні такого набору ознак Z, знайденого в класі F допустимих перетворень початкових показників $X^{(1)}$, $X^{(2)}$, …, $X^{(P)}$, що

$$I_{P'}(Z(x)) = \max_{z\in F}\{I_{P'}(Z(X))\}, \tag{2}$$

Варіант конкретизації заходів інформативності призводить до конкретного виду зниження розмірності: до методу головних компонент, факторному аналізу, угруповання параметрів та інше.

Відповідно до поставленої задачі зупинимося на методі головних компонент.

Головні компоненти являють собою ортогональні лінійні перетворення векторної випадкової величини X, такі, що перша з них має найбільшу дисперсію, дисперсія зменшується з ростом номера змінної, так що P-а має мінімальну дисперсію [3]. Вибираються P' перших головних компонент за таким правилом:

$$\begin{gathered} Z^{(i)}(X) = C_{j1}(X^{(1)} - M^{(1)}) + ... + C_{jP}(X^{(P)} - M^{(P)}), \quad i = \overline{1,P'}, \\ \sum_{v=1}^{P} C_{jv}^2 = 1, \quad j = \overline{1,P}, \\ \sum_{v=1}^{P} C_{jv}, C_{kv} = 0, \quad j,k = \overline{1,P}, j \neq k. \end{gathered}, \tag{3}$$

де $M^{(v)} = M[X^{(v)}]$ – математичне очікування ознаки $X^{(v)}$.

В якості міри інформативності P'-мірної системи показників $Z^{(1)}, Z^{(2)},..., Z^{(P')}$ прийнято вираз:

$$I_{P'}(Z(X)) = \frac{DZ^{(1)} + ... + DZ^{(P')}}{DX^{(1)} + ... + DX^{(P)}}, \tag{4}$$

де D – операція обчислення дисперсії.

Використання головних компонент виявляється найбільш природним і плідним в ситуаціях, в яких всі компоненти $X^{(1)}$, $X^{(2)}$, …, $X^{(P)}$ досліджуваного вектора **X** мають загальну фізичну природу і відповідно виміряні в одних і тих же одиницях. Якщо ж різні ознаки $X^{(1)}$, $X^{(2)}$, …, $X^{(P)}$ виміряні в різних одиницях, то результати дослідження будуть істотно залежати від вибору масштабу і природи одиниць виміру. Цей випадок характерний для задач селекції, тому тут доцільно попередньо переходити до допоміжних безрозмірних ознак

$$X_v^{*(i)} = \frac{X_v^{(i)} - M^{(i)}}{\sqrt{D_{ii}}}, \quad i = \overline{1,P}, \quad v = \overline{1,n}. \tag{5}$$

Для знаходження головних компонент необхідно за наявною статистичною сукупністю знайти $M^{(i)}, D_{ii}$, побудувати кореляційну матрицю **V**.

Тоді, з урахуванням (3) та (5) можна записати вираз для першої головної компоненти

$$Z_1 = \sum_{i=1}^{P} C_{1i}\overline{X}_i = \overline{C_1'}\,\overline{X} \tag{6}$$

Ясно, що $M[Z_1] = 0$ та $Z_1 = M[Z_1^2] = \sum_{i=1}^{P}\sum_{j=1}^{P} C_{1i}C_{1j}M\left[\overline{X}_i\overline{X}_j\right] = \overline{C_1'}\,\mathbf{V}\,\overline{C_1 = D_1}$.

Вектор коефіцієнтів $\overline{C_1}$, вибраний таким чином, щоб дисперсія Z_1 мала максимальне значення при умові

$$Z_1 = \sum_{i=1}^{P} C_{1i}^2 = \overline{C_1'}\,\overline{C_1} = 1 \tag{7}$$

Таким чином, виникає проблема максимізації при наявності обмежень. Вирішимо її методом невизначених множників Лагранжа [3, 4]. Знайдемо $\overline{C_1}$, що доставляє максимум виразу

$\overline{C_1'}\,\mathbf{V}\,\overline{C_1}-\lambda_1(\overline{C_1'}\,\overline{C_1}-1)$, де λ_1 – множник Лагранжа. Взявши похідну по $\overline{C_1}$, і прирівнявши її до 0, отримуємо рівняння

$$(\mathbf{V}-\lambda_1\mathbf{I})\overline{C_1}=0 \tag{8}$$

де **I** – одинична матриця.

Так як цікавить розв'язання, коли $\overline{C_1}\neq 0$, то отримаємо

$$|\mathbf{V}-\lambda_1\mathbf{I}|=0 \tag{9}$$

Отже, λ_1 – власне число, а $\overline{C_1}$ – відповідний власний вектор. Перепишемо (8) у виді $\mathbf{V}\overline{C_1}=\lambda_1\overline{C_1}$ і помножимо зліва на $\overline{C_1'}$ отримаємо $\mathbf{V}\overline{C_1}=\lambda_1\,\overline{C_1'}\,\overline{C_1}=\lambda_1$, де ліва частина згідно (6) є var Z_1, а λ_1 є максимальне власне число матриці **V**. Щоб знайти другу головну компоненту $Z_2=\overline{C_2'}X$ вимагатимемо виконання двох умов:

– умови нормування

$$\overline{C_2'}\,\overline{C_2}=1 \tag{10}$$

– умови ортогональності

$$\overline{C_1'}\,\overline{C_2}=0 \tag{11}$$

Вектор $\overline{C_2}$ визначається тепер так, щоб var (Z_2), була максимальна при виконанні двох зазначених умов. Ця задача вимагає використання двох множників Лагранжа λ_2 і β. При цьому необхідно максимізувати вираз

$$\overline{C_2'}\,\mathbf{V}\,\overline{C_2}-\lambda_2\left(\overline{C_2'}\,\overline{C_2}-1\right)-\beta\left(\overline{C_1'}\,\overline{C_2}-0\right) \tag{12}$$

Взявши похідну від (12) і прирівнявши її 0, знаходимо відповідно до умови ортогональності (11), що $\beta = 0$. А, в силу умови нормування (10), отримуємо, що λ_2 є друге за величиною власне число матриці **V**, λ_2 = var (Z_2), а $\overline{C_2}$ – відповідний власний вектор. Процес повторюється до тих пір, поки всі власні числа і власні вектори не опиняться дисперсіями і коефіцієнтами лінійних комбінацій головних компонент. На цьому закінчена перша частина роботи – знайдені лінійні некорельовані комбінації ознак рослин X і Y, що схрещується, які мають найбільший розкид (найбільші дисперсії). Позначимо їх, відповідно,

$$\begin{aligned}\overline{X}&=(Z_1^{(1)},Z_2^{(1)},...,Z_P^{(1)})'=(X_1,X_2,...,X_P)',\\ \overline{Y}&=(Z_1^{(2)},Z_2^{(2)},...,Z_P^{(2)})'=(Y_1,Y_2,...,Y_P)'.\end{aligned} \tag{13}$$

Нехай коваріаційні (кореляційні) матриці для $\overline{X}$ і $\overline{Y}$ будуть відповідно

$$V_{11}=M[\overline{X},\overline{X}'],\quad V_{22}=M[\overline{Y},\overline{Y}'],$$

а матриця взаємної кореляції буде $V_{12}=M[\overline{X},\overline{Y}']$. Цікавить питання знаходження головних компонент множини $\overline{X}$ і множини $\overline{Y}$, які найбільш корелюються, що дає більш економний опис зв'язку $\overline{X}$ та $\overline{Y}$, ніж задається матрицею V_{12} [5, 6].

З урахуванням того, що

$$\begin{aligned}\lambda_i^{(x)}&=\overline{\alpha_i'}\,V_{11}\,\overline{\alpha_1}=D_i^{(x)},\\ \lambda_i^{(y)}&=\overline{\beta_i'}\,V_{22}\,\overline{\beta_1}=D_i^{(y)},\\ \overline{\alpha_i'}\,\overline{\alpha_1}&=1,\quad \overline{\beta_i'}\,\overline{\beta_1}=1\\ D^{(x)}&=\overline{\alpha'}\,V_{11}\,\overline{\alpha}=1,\\ D^{(y)}&=\overline{\beta'}\,V_{22}\,\overline{\beta}=1,\end{aligned} \tag{14}$$

коефіцієнт кореляції між $\overline{X}$ та $\overline{Y}$ буде

$$\rho = \overline{\alpha'}\, V_{12}\, \overline{\beta}\,. \tag{15}$$

Визначимо вектори $\overline{\alpha}$ і $\overline{\beta}$ при яких (15) досягає максимуму при виконанні умов (14).

Запишемо для цього функцію Лагранжа

$$\psi = \overline{\alpha'}\, V_{12}\, \overline{\beta} - v_1(\overline{\alpha'}\, V_{11}\, \overline{\alpha} - 1) - v_2(\overline{\beta'}\, V_{22}\, \overline{\beta} - 1), \tag{16}$$

$$\begin{aligned} \frac{\partial \psi}{\partial \overline{\alpha}} &= V_{12}\, \overline{\beta} - 2v_1\, V_{11}\, \overline{\alpha} = 0, \\ \frac{\partial \psi}{\partial \overline{\beta}} &= V'_{12}\, \overline{\alpha} - 2v_2\, V_{22}\, \overline{\beta} = 0 \end{aligned} \tag{17}$$

Помноживши перше рівняння на $\overline{\alpha'}$, а друге на $\overline{\beta'}$, отримаємо:

$$\begin{aligned} 2v_1\, \overline{\alpha'}\, V_{11}\, \overline{\alpha} &= \overline{\alpha'}\, V_{12}\, \overline{\beta}, \\ 2v_2\, \overline{\beta'}\, V_{22}\, \overline{\beta} &= \overline{\beta'}\, V'_{12}\, \overline{\alpha}. \end{aligned}$$

З урахуванням обмежень $\overline{\alpha'}\, V_{11}\, \overline{\alpha} = 1$, $\overline{\beta'}\, V_{22}\, \overline{\beta} = 1$ та $\overline{\beta'}\, V'_{12}\, \overline{\alpha} = \overline{\alpha'}\, V_{12}\, \overline{\beta} = \rho_1$ тягнуть рівність $2v_1 = 2v_2 = \lambda_1$. Ця величина і є максимальне значення ρ. З першого рівняння (17) маємо:

$$\overline{\alpha} = \frac{1}{\lambda} V_{11}^{-1}\, V_{12}\, \overline{\beta}. \tag{18}$$

Підставимо його в друге рівняння (17), отримаємо

$$\begin{aligned} (V'_{12}\, V_{11}^{-1}\, V_{22} - \lambda_1\, V_{22})\overline{\beta} &= 0, \\ (V_{22}^{-1}\, V'_{12}\, V_{11}^{-1}\, V_{12} - \lambda_1\, \mathbf{I})\overline{\beta} &= 0, \end{aligned}$$

виходячи з того, що цікавить ненульове рішення, то

$$\left| V_{22}^{-1}\, V'_{12}\, V_{11}^{-1}\, V_{12} - \lambda_1\, \mathbf{I} \right| = 0.$$

Найбільше власне значення матриці $\left| V_{22}^{-1}\, V'_{12}\, V_{11}^{-1}\, V_{12} \right|$ є ρ_1, а $\overline{\beta}_1$-власний вектор. Аналогічно отримуємо $\left| V_{11}^{-1}\, V'_{12}\, V_{22}^{-1}\, V_{12} - \lambda_1\, \mathbf{I} \right| \overline{\alpha} = 0$, той же результат для ρ_1 і власний вектор $\overline{\alpha}_1$. Тоді перша пара канонічних змінних

$$\begin{aligned} \overline{X}_1 &= \overline{\alpha'}_1\, \overline{X}, \\ \overline{Y}_1 &= \overline{\beta'}_1\, \overline{Y}. \end{aligned}$$

Їх коефіцієнт кореляції $\rho = \lambda_1$.

Друга пара канонічних змінних

$$\begin{aligned} \overline{X}_2 &= \overline{\alpha'}_2\, \overline{X}, \\ \overline{Y}_2 &= \overline{\beta'}_2\, \overline{Y}. \end{aligned}$$

вибирається з умови максимуму їх коефіцієнта кореляції $\overline{\alpha'_2}\, V_{12}\, \overline{\beta}_2$ з урахуванням умови одиничності дисперсій і умов ортогональності

$$\operatorname{cov}(\overline{X}_1\ \overline{X}_2) = \operatorname{cov}(\overline{Y}_1\ \overline{Y}_2) = 0\,,$$

тобто

$$\overline{\alpha'_1}\, V_{11}\, \overline{\alpha}_2 = \overline{\beta'_1}\, V'_{22}\, \overline{\beta}_2 = \overline{\beta'_1}\, V_{12}\, \overline{\alpha}_2 = 0.$$

Вектори $\overline{\alpha}_2$ та $\overline{\beta}_2$, відповідні максимальному значенню ρ, виходять аналогічним методом – розв'язанням функції Лагранжа:

$$\psi = \overline{\alpha_2'}\, V_{12}\, \overline{\beta}_2 - v_1(\overline{\alpha_2'}\, V_{11}\, \overline{\alpha}_2 - 1) - v_2(\overline{\beta_2'}\, V_{22}\, \overline{\beta}_2 - 1) + v_3 \overline{\alpha_1'}\, V_{11}\, \overline{\alpha}_2 + v_4 \overline{\beta_1'}\, V_{22}\, \overline{\beta}_2 + v_5 \overline{\beta_1'}\, V_{12}\, \overline{\alpha}_2.$$

З урахуванням умови ортогональності після взяття похідних отримуємо $v_3 = v_4 = v_5 = 0$, і, з урахуванням нормування, $2v_1 = 2v_2 = \lambda_2$. Друге найбільше значення ρ_2 є друге за величиною власне число матриці $V_{11}^{-1} V_{12}' V_{22}^{-1} V_{12}$, а вектор $\overline{\beta}_2$ є власний вектор цієї матриці, (аналогічно знаходиться вектор $\overline{\alpha}_2$).

Продовжуючи цю процедуру можемо отримати набір коефіцієнтів кореляції між парами канонічних змінних $(\overline{X}_i, \overline{Y}_i)$ і, аналізуючи значущі з точки зору абсолютних значень ρ_i, підбирати поєднання параметрів примірників рослин, що схрещується, згідно поставленого критерію (на врожайність, на якість, на стійкість, на агротехнічні ознаки).

Висновки

Таким чином, пропонований метод цілеспрямованого пошуку пар для схрещування зводиться до двох етапів: перший – пошук головних компонент параметрів рослин, що схрещується, і визначення їх кількості, виходячи з максимуму сумарної дисперсії, що визначаються головними компонентами; другий – перебування канонічних кореляцій для знайдених пар головних компонент і вторинних векторів, елементами яких є знайдені головні компоненти. У цьому основний зміст пропонованого алгоритму спрямованого пошуку. Перевірка значущості знайдених канонічних коефіцієнтів кореляції може бути проведена за узагальненим критерієм χ^2 Пірсона. Експериментальна перевірка для конкретних видів рослин може дати: повне підтвердження запропонованого методу, часткове підтвердження або повне заперечення. Швидше за все найбільш ймовірним є другий результат в силу стохастичності початкового матеріалу.

Список використаної літератури

1. Использование генетических методов в селекции. 2021. URL: https://spravochnick.ru/biologiya/genetika_kak_nauka/ispolzovanie_geneticheskih_metodov_v_selekcii/ (дата звернення 23.04.21).
2. Вавилов Н. И. Теоретические основы селекции / Н.И. Вавилов. М.: Наука, 1987. 512 с.
3. Иберла К. Факторный анализ. / М. Иберла. М.: Статистика, 1980. 398 с.
4. Андерсон Т. Введение в многомерный статистический анализ / Т. Андерсон. М.: Физматгиз, 1963. 500 с.
5. Экономико-математические методы и прикладные модели / В.В. Федосеев, А.Н. Гармаш, Д.М. Дайитбегов и др.; Ред. В.В. Федосеева. М.: ЮНИТИ, 1999. 391 с.
6. Димова Г.О. Методи і моделі упорядкування експериментальної інформації для ідентифікації і прогнозування стану безперервних процесів: монографія. / Ганна Олегівна Димова. Херсон: Видавництво ФОП Вишемирський В.С., 2020. 176 с.

References

1. Ispol'zovaniye geneticheskikh metodov v selektsii. [The use of genetic methods in breeding] 2021. URL: https://spravochnick.ru/biologiya/genetika_kak_nauka/ispolzovanie_geneticheskih_metodov_v_selekcii/ (reference date 04.23.21).
2. Vavilov N.I. Teoreticheskiye osnovy selektsii [Theoretical Foundations of Breeding] Moscow: Nauka, 1987.512 p.
3. Iberla K. Faktornyy analiz [Factor analysis] Moscow: Statistics, 1980.398 p.
4. Anderson T. Vvedeniye v mnogomernyy statisticheskiy analiz [Introduction to multivariate statistical analysis] Moscow: Fizmatgiz, 1963.500 p.
5. Fedoseev V.V., Garmash A.N., Dayitbegov D.M. and others. Ekonomiko-matematicheskiye metody i prikladnyye modeli [Economic and mathematical methods and applied models] Ed. V.V. Fedoseev. Moscow: YUNITI, 1999.391 p.
6. Dymova H.O. Metody i modeli uporyadkuvannya eksperymental'noyi informatsiyi dlya identyfikatsiyi i prohnozuvannya stanu bezperervnykh protsesiv: monohrafiya [Methods and models for ordering experimental information for identifying and predicting the state of continuous processes] Kherson: Publishing house FOP Vyshemyrskyy V.S., 2020. 176 p.