

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL AND COMPUTER MODELLING

УДК 528.85

Шама Є. О.

Аспірант кафедри радіотехніки та телекомунікацій Запорізького національного технічного університету, Запоріжжя, Україна

ЗМЕНШЕННЯ РОЗМІРНОСТІ ОЗНАКОВОГО ПРОСТОРУ РОСЛИННИХ ОБ'ЄКТІВ ЗА ДОПОМОГОЮ ФАКТОРНОГО АНАЛІЗУ

У статті показана можливість зменшення кількості інформативних ознак для багатоспектральних даних рослинного походження за допомогою методів факторного аналізу. Для дослідження були використані реальні коефіцієнти відбиття кукурудзи, мишію та амброзії. Перед використанням метода факторного аналізу була виконана процедура перевірки відповідності початкового набору даних до нормального закону розподілу. Факторний аналіз проводився окремо для трьох рослин: кукурудзи, мишію та амброзії. Для кожної рослини встановлені розміри нового простору, а також надана якісна оцінка кореляції між елементами нового і початкового простору даних. Перехід до нового ознакового простору виконувався за умови, що рівень інформативності підтримувався на рівні не менше 80% по відношенню до початкового набору даних. Для кожного виду рослини була проведена процедура перевірки правильності вибору кількості ознак за допомогою розрахунку спільностей. Отримані результати можуть бути використані при побудові класифікатора рослинних об'єктів.

Ключові слова: факторний аналіз, ознаковий простір, спектральні коефіцієнти яскравості, ознаки, класифікатор.

НОМЕНКЛАТУРА

a_{ik} – факторні навантаження i -го значення k -го фактору;

F_k – k -й загальний фактор;

h_i^2 – спільність i -го значення;

N – кількість вимірювань, що зроблені для окремої рослини за допомогою спектрометра;

P – кількість факторів;

S_i^2 – специфічність i -го значення;

S_j – стандартне відхилення j -го значення;

$x_{\bullet j}$ – середнє значення j -ї ознаки;

X_j – j -та спостережувана змінна;

z_{ij} – компоненти матриці стандартизованих початкових даних;

U_j – j -та складова специфічних факторів;

λ_i – i -та довжина хвилі;

μ_i – i -й вимірюваний спектральний коефіцієнт відбиття;

$\mu_{i,\max}$ – i -й максимальний вимірюваний спектральний коефіцієнт відбиття;

$\mu_{i,\min}$ – i -й мінімальний вимірюваний спектральний коефіцієнт відбиття;

ω_i – i -й нормований спектральний коефіцієнт відбиття.

ВСТУП

Відомо, що велика розмірність ознакового простору становить суттєву проблему для багатьох процедур розпізнавання образів, зокрема рослинних об'єктів.

Об'єктом дослідження є процес зменшення розмірності простору ознак для побудови розпізнавальних моделей.

На сьогоднішній день зростає кількість теоретичних робіт по методам зменшення розмірності ознакового простору для класифікації образів [1]. Деякі з цих методів прагнуть сформувати нові ознаки на основі лінійних комбінацій старих ознак, а інші прагнуть створити меншу підмножину початкових ознак. Основна проблема цієї теорії полягає в тому, що поділ розпізнавання образів на виділення ознак, а потім класифікацію теоретично є штучним. Повністю оптимальний видільник ознак є не чим іншим, як оптимальним класифікатором. А тому, вирішивши задачу зменшення інформативних ознак, при збереженні необхідного рівня інформативності, можна спростити задачу побудови класифікатора рослинних об'єктів.

Більшість запропонованих методів зменшення розмірності [1] забезпечують функціональне відображення, так що можливо знайти відображення довільного вектора ознак [2]. Класичними процедурами зменшення розмірності є аналіз головних компонент і факторний аналіз

[3], обидва з яких зменшують розмірність шляхом формування лінійних комбінацій ознак. Метою аналізу головних компонент (розкладання Кархунена-Лоева) є знаходження зображень у просторі меншої розмірності, яке враховує дисперсію ознак. Метою факторного аналізу є знаходження зображень у просторі меншої розмірності, яке враховує кореляцію між ознаками. Різні алгоритми факторного аналізу об'єднані тим, що в усіх них відбувається перехід до нового базису в початковому n -мірному просторі, тому у статті буде використаний факторний аналіз, так як він є загальним методом зменшення розмірності серед методів головних компонентів та факторного аналізу.

Предметом дослідження є процедура зменшення розмірності ознакового простору коефіцієнтів відбиття рослинних об'єктів за допомогою методів факторного аналізу. При цьому, початкова (перед проведенням процедури зменшення ознакового простору) розмірність коефіцієнтів відбиття рослинних об'єктів, складає 256, яка визначається роздільною здатністю вимірювального приладу – спектрометра.

Мета роботи полягає у дослідженні можливості зменшення кількості інформаційних ознак за допомогою використання методів факторного аналізу.

1 ПОСТАНОВКА ЗАДАЧІ

Нехай ми маємо масив експериментально отриманих даних коефіцієнтів відбиття рослинних об'єктів X_j . Розмірність масиву коефіцієнтів відбиття рослинних об'єктів – $256 \times N$, де 256 – розмірність ознакового простору (визначається роздільною здатністю спектрометра). Тоді, задача зменшення ознакового простору рослинних об'єктів, полягає у тому, щоб отримати масив меншої розмірності даних для відповідної рослини, за допомогою якого можна зробити процедуру розпізнавання із заздалегідь встановленим рівнем вірогідності. Тобто необхідно згенерувати вихідний масив даних F_k , який був би меншим за розміром і за допомогою якого можливо було б зробити процедуру розпізнавання рослини за видом. Так, якщо вихідний масив можливо описати

за допомогою лінійного перетворення $F_k = \sum_{j=1}^m a_{jk} \cdot X_j$,

то вирішувана задача в рамках цієї статті полягає в здобутті нового масиву, у якого $m \ll N$ і $m \rightarrow \min$.

2 ОГЛЯД ЛІТЕРАТУРИ

Найбільш відомими методами розпізнавання є метод потенціальних функцій [1, 7], дискримінантний аналіз [1, 2, 7], метод граничних спрощень [8], регресійний аналіз [2], метод групового урахування аргументів [7], нейромережеві методи [1, 2]. Порівняльний аналіз основних методів розпізнавання – кореляційних, ознакових (статистичних і детермінованих), нейромережевих [2] – дозволяє зробити висновок, що метод факторного аналізу є найбільш ефективним для вирішення завдань розпізнавання [9]. Крім того, факторні методи забезпечують більш високу стійкість роботи системи розпізнавання в умовах перешкод і дозволяють об'єднати процедуру розпізнавання і процедуру зменшення ознак.

Розпізнавання рослинних об'єктів за допомогою факторного аналізу розглянуто у ряді робіт [3, 5, 9]. Але в цих роботах класифікація рослинних об'єктів проводилась лише на два класи (корисну культуру і бур'яни). Проте на практиці доцільною є класифікація на більшу кількість класів, що обумовлюється необхідністю використання різних гербіцидів для різних видів бур'янів. Більша кількість класів обумовлена тим, що для боротьби з бур'янами при хімічній обробці застосовують різні види гербіцидів. Необхідність розпізнавання рослинних об'єктів на три класи визначається також алгоритмом роботи інтелектуального обприскувача для обробки просяних культур в рослинництві. Також процедура розпізнавання в моделях на основі дискримінантного аналізу [1, 6], нейронної мережі [2, 4, 8], регресійного аналізу [7], як правило, характеризується великою надлишковістю і в них висувуються високі вимоги до обчислювальних ресурсів. Для того, щоб виключити недоліки цих методів, доцільно застосувати процедуру факторного аналізу для зменшення надлишковості даних моделі при розпізнаванні рослинних об'єктів, що скоротить час навчання та суттєво збільшить швидкість прийняття рішень і дозволить зменшити вимоги до обчислювальних ресурсів.

3 МАТЕРІАЛИ І МЕТОДИ

Факторний аналіз дозволяє розв'язати дві важливі проблеми, які виникають під час дослідження: компактно описати об'єкт виміру і в той же час всебічно. За допомогою факторного аналізу можливе виявлення прихованих змінних факторів, що відповідають за наявність лінійних статистичних кореляцій між спостережуваними змінними [4].

При проведенні факторного аналізу, в один чинник об'єднуються змінні, що сильно корелюють між собою, як наслідок відбувається перерозподіл дисперсії між компонентами і виходить максимально проста і наявна структура факторів. Після об'єднання корельованість компонентів, всередині кожного фактору між собою, буде вищим, ніж їх корельованість з компонентами з інших факторів. Ця процедура також дозволяє виділити латентні змінні, що може бути корисним при побудові класифікатора рослинних об'єктів.

Таким чином можна виділити дві мети факторного аналізу:

- визначення взаємозв'язків між змінними (класифікація змінних);
- зменшення кількості змінних необхідних для опису даних.

При факторному аналізі, коли необхідно подати початкові змінні у вигляді лінійної комбінації факторів F , які розраховані так, щоб найкращим способом (з мінімальною погрішністю) подати X , можна показати [10]:

$$X_j = \sum_{k=1}^p a_{jk} \cdot F_k + U_j, \quad k=1,2,\dots,p; \quad j=1,2,\dots,m. \quad (1)$$

Одним з найбільш поширених методів пошуку факторів є метод головних компонент (саме цей метод буде використовувати в межах факторного аналізу). Його основна відмінність від факторного аналізу полягає в тому,

що головні компоненти F_k пов'язані із спостережуваними змінними X_j лінійними функціями перетворення:

$$X_j = \sum_{k=1}^p a_{jk} \cdot F_k \text{ і } F_k = \sum_{j=1}^m a_{jk} \cdot X_j. \quad (2)$$

Суть цього методу полягає в заміні корельованих компонентів некорельованими чинниками. Іншою важливою характеристикою методу є можливість обмежитися найбільш інформативними головними компонентами і виключити інші з аналізу, що спрощує інтерпретацію результатів. Цінність методу головних компонентів також в тому, що він – єдиний математично обґрунтований метод факторного аналізу [5]. Але слід зауважити, що за твердженням деяких дослідників, метод головних компонентів не є методом факторного аналізу, оскільки не розщеплює дисперсію індикаторів на загальну і унікальну [6].

Метод головних компонент є простішим в розрахунках і інтерпретації, але одна з головних труднощів його використання – необхідність перетворення початкових даних, які представлені в різних одиницях виміру, в величини, які можливо співставити. Традиційним методом перетворення, але не обов'язковим, є нормування по стандартним відхиленням, коли матриця Z стандартизованих початкових даних визначається за формулою:

$$z_{ij} = \frac{x_{ij} - x_{\cdot j}}{S_j}, j=1,2,\dots,m; i=1,2,\dots,n. \quad (3)$$

Також слід зауважити, що при проведенні факторного аналізу за допомогою методу головних компонент може використовуватись така характеристика, як спільність, яка визначається формулою:

$$h_i^2 = \sum_{k=1}^M a_{ik}^2 + S_i^2, \quad (4)$$

Тобто спільність є дисперсією змінних, обумовленою факторами, а специфічність являє собою частину дисперсії, яка обумовлена випадковими помилками або змінними, неврахованими в моделі. Відповідно до постановки завдання необхідно шукати такі фактори, при яких сумарна спільність є максимальною, а специфічність – мінімальною.

У роботі [7] виділяють такі умови проведення факторного аналізу: всі ознаки повинні бути кількісними; кількість спостережень має бути не менше як у два рази більшою кількості змінних; вибірка має бути однорідною; факторний аналіз здійснюється по корелюючими змінними.

4 ЕКСПЕРИМЕНТИ

Виміри спектральних коефіцієнтів відбиття проводилися влітку 2013 року на полі ТОВ «Агрофірма «Матвіївка» (Запорізька область, Україна). В якості корисної культури була обрана кукурудза. Також на полі зустрічались такі бур'яни: амброзія, берізка, мишій, пирій, пастуша сумка, суріпиця та ін. Серед бур'янів, для дослідження, були відібрані мишій та амброзія, оскільки вони склали переважну долю бур'янів на полі. Вимірювання проводилися максимально наближеним до реальних польових робіт на самохідній установці: швидкість руху 0,5–

1 м/с, відстань об'єктива до рослини – приблизно 0,3 м. Переміщення об'єктива приймача спектрометра від однієї рослини до іншої в продовж рядка проводилося рівномірно, без ривків. Вимірювання проводилися впродовж проходу міжряддя як в одному напрямку так і в іншому. Пристрій, за допомогою якого проводилися вимірювання спектральних коефіцієнтів відбиття, містив спектрометр Red Tide650 з волоконно-оптичним кабелем P200 – 2 – UV – VIS і лінзою 74 – VIS (фірма Ocean Optics) та веб-камеру A4Tech PK-838G.

Вимірювання спектральних характеристик рослин проводилися за допомогою приладу, який автоматично з періодом 1 с зберігав поточну інформацію (коефіцієнт відбиття і координати місця) і фотографію зовнішнього вигляду рослини.

З усіх експериментальних даних, для подальшого дослідження було відібрано 2268 спектральних коефіцієнтів відбиття рослин. Критерієм відбору була можливість візуального визначення виду рослин по зображеннях і формі спектральної кривої. Назва і кількість рослин, які були відібрані для подальшого дослідження, наведені в таблиці 1.

Таблиця 1 – Видовий склад досліджуваних рослин

Вид рослини	Латинська назва	Кількість екземплярів
Кукурудза	<i>Zéa máys</i>	699
Амброзія	<i>Ambrósia artemisiifólia</i>	882
Мишій	<i>Setaria viridis</i>	687
Усього		2268

РЕЗУЛЬТАТИ

За допомогою спектрометра експериментально були отримані спектральні коефіцієнти відбиття рослин, які були потім експериментально оброблені. Було отримано три матриці: матриця коефіцієнтів відбиття для кукурудзи – 699×256, матриця коефіцієнтів відбиття для мишію – 687×256 і матриця коефіцієнтів відбиття для амброзії – 882×256. Наступним етапом було нормування цих даних. Кожна з трьох матриць піддавалася нормуванню за лінійним законом:

$$\omega_i = \frac{\mu_i - \mu_{i,\min}}{\mu_{i,\max} - \mu_{i,\min}}. \quad (5)$$

Перед проведенням факторного аналізу необхідно перевірити генеральну сукупність даних для кожної рослини на відповідність нормальному розподілу.

Нормальність розподілу даних для кукурудзи, мишію і амброзії була перевірена у статистичному пакеті програм. Для перевірки розподілу на нормальність використовувались нормовані значення мишію. Результати перевірки наведені на рис. 1 і в табл. 2.

Таким чином, аналіз досліджуваного розподілу вибірок визначив його наближеність до нормального, на що вказують як вид гістограми, так і значення коефіцієнтів асиметрії та ексцесу: вони одного порядку по відношенню до початкових значень вибірок, а також мають малі значення стандартних помилок. Для усіх видів аналізу статистично значимими вважали відмінності при $P < 0,05$.

Набір початкових ознак для кукурудзи та амброзії також піддавався аналогічній процедурі перевірки на нормальність і також було встановлено наближеність роз-

поділів цих рослин до нормального закону (що не дивно, бо природа походження і умови виміру коефіцієнтів якравості для усіх трьох рослин є однаковими).

Наступним етапом було проведення факторного аналізу. При проведенні факторного аналізу в програмному пакеті були встановлені наступні параметри: метод виділення факторів – компонентний аналіз; максимальна кількість факторів – 256; мінімальне власне число – 0.

Факторний аналіз будемо використовувати окремо для кожного виду рослинних об'єктів: корисної культури (кукурудза), однодольного бур'яну (мишії), дводольного бур'яну (амброзія). Умовою переходу від початкового *n*-мірного простору до нового ознакового простору, буде

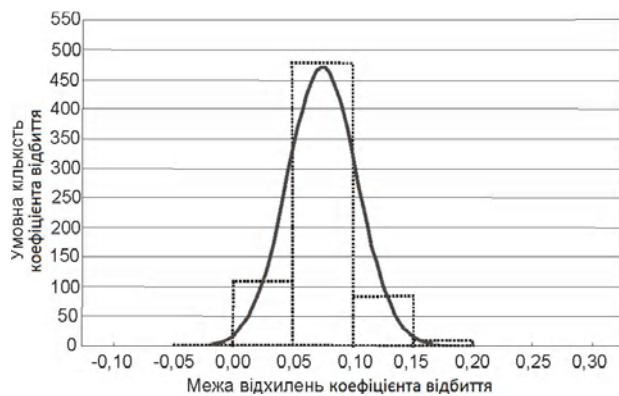


Рисунок 1 – Гістограма усередненої вибіркової сукупності значень мишію. Суцільна лінія – крива нормального закону розподілу

рівень інформативності нового простору, який повинен бути не менше 80% по відношенню до початкового.

Рівень інформативності факторного ознакового простору встановимо рівним 80% по відношенню до початкового. Це є достатнім рівнем інформативності при переході від одного ознакового простору до іншого [8]. Для визначення кількості факторів, які можна залишити і при цьому відповідати встановленому рівню інформативності визначимо вклад головних компонент для кожної рослини.

На таблицях 3 та 4 в перших стовбцях наведені оцінки власних чисел, в третьому стовбці – накопичені значення власних чисел, у другому і у четвертому стовбцях – відносний вклад кожної головної компоненти та накопичений відносний вклад відповідно.

У першому наближенні накопичений відносний вклад можна вважати рівнем інформативності нового ознакового простору по відношенню до початкового (для більш точного значення рівня інформативності необхідно зробити перерахунок через власні числа з урахуванням довірчих інтервалів) [9]. Як бачимо, для досягнення встановленого рівня інформативності, для кукурудзи можна залишити 3 фактора, для мишію – 8 факторів і для амброзії – 2 фактора.

Для інтерпретації нових ознак необхідно провести аналіз матриці факторних навантажень з урахуванням нової кількості факторів для кожної рослини.

5 ОБГОВОРЕННЯ

Як бачимо з таблиць 5 і 6, для кукурудзи перший фактор найбільш корельований з початковими ознаками, що відповідають довжинам хвиль 353–464 нм, 631–704 нм і

Таблиця 2 – Результати розрахунку критеріїв асиметрії та ексцесу для мишію

Змінна	Середнє значення	Мінімум	Максимум	Станд. відхилення	Асиметрія	Похибка асиметрії	Ексцес	Похибка ексцесу
1	0,083706	0	0,600420	0,071435	1,47213	0,093251	5,46625	0,186233
2	0,015076	0	0,124923	0,018389	1,58612	0,093251	3,48430	0,186233
3	0,013580	0	0,153521	0,018177	2,09398	0,093251	7,26450	0,186233
4	0,057173	0	0,203566	0,025330	1,11122	0,093251	3,16737	0,186233
5	0,054378	0	0,243602	0,025989	1,15261	0,093251	4,32015	0,186233
6	0,059009	0	0,225000	0,027683	1,38278	0,093251	5,47856	0,186233
7	0,054660	0	0,204996	0,026185	0,82713	0,093251	1,94301	0,186233

Таблиця 3 – Вклад головних компонент в сумарну дисперсію початкових ознак для кукурудзи і для мишію

№ фактору	Власні значення для кукурудзи				Власні значення для мишію			
	Значення	% від загального	Накопичене значення	% накопичення	Значення	% від загального	Накопичене значення	% накопичення
1	163,3985	63,827	163,3985	63,827	101,1679	39,5187	101,1679	39,518
2	29,9047	11,681	193,3032	75,509	70,7136	27,6224	171,8815	67,141
3	13,4457	5,252	206,7489	80,761	15,0487	5,8784	186,9302	73,019
4	2,9825	1,165	209,7314	81,926	9,94743	3,7009	196,4046	76,720
5	1,5078	0,588	211,1577	82,515	4,8562	1,8969	201,2607	78,617
6	0,9185	0,358	212,1577	82,874	1,5239	0,5952	202,7847	79,212
7	0,8381	0,327	212,9958	83,201	1,4546	0,5682	204,2393	79,781
8	0,7767	0,303	213,7725	83,504	1,0522	0,4110	205,2914	80,192
9	0,7139	0,278	214,4864	83,783	1,0216	0,3990	206,3130	80,591
10	0,6942	0,271	215,1806	84,057	0,9499	0,3710	207,2629	80,962
11	0,6580	0,257	215,8387	84,312	0,8506	0,3322	208,1135	81,294
12	0,6277	0,245	216,4664	84,557	0,8196	0,3201	208,9331	81,614
13	0,5955	0,232	217,0618	84,789	0,7968	0,3112	209,7298	81,925
14	0,5709	0,223	217,6327	85,012	0,7567	0,29558	210,4865	82,221
15	0,5537	0,216	218,1864	85,229	0,7094	0,2771	211,1959	82,498
16	0,5488	0,214	218,7352	85,443	0,7006	0,2773	211,8965	82,772
17	0,5163	0,201	219,2515	85,645	0,6876	0,2685	212,5841	83,040
18	0,5048	0,197	219,7563	85,842	0,6605	0,2580	213,2446	83,298
19	0,4910	0,191	220,2474	86,034	0,6506	0,2541	213,8951	83,552

Таблиця 4 – Вклад головних компонент в сумарну дисперсію початкових ознак для амброзії

№ фактору	Власні значення для амброзії			
	Значення	% від загального	Накопичене значення	% накопичення
1	145,327	56,76864	145,3277	56,7686
2	79,2253	30,94740	224,5530	87,7160
3	13,0125	5,08299	237,5655	92,7990
4	6,7203	2,62510	244,2858	95,4241
5	3,1865	1,24471	247,4722	96,6688
6	1,4380	0,56172	248,9103	97,2306
7	0,9799	0,38276	249,8901	97,6133
8	0,7389	0,28863	250,6290	97,9020
9	0,6168	0,24094	251,2458	98,1429
10	0,5851	0,22856	251,8309	98,3715
11	0,3474	0,13571	252,1783	98,5072
12	0,2324	0,09077	252,4107	98,5979
13	0,1990	0,07772	252,6097	98,6757
14	0,1852	0,07233	252,7948	98,7480
15	0,17764	0,06892	252,9713	98,8169
16	0,1617	0,06318	253,1330	98,8801
17	0,1588	0,06203	253,2918	98,9421
18	0,1416	0,05530	253,4334	98,9974
19	0,1335	0,05215	253,5669	99,0496
20	0,1194	0,04475	253,6862	99,0962
21	0,1146	0,06318	2553,8008	99,1409

Таблиця 5 – Факторні навантаження для кукурудзи та для амброзії

Змінні	Факторні навантаження (для кукурудзи)			Факторні навантаження (для амброзії)	
	Фактор 1	Фактор 2	Фактор 3	Фактор 1	Фактор 2
1	0,3154	-0,00652	0,13011	-0,5882	0,0373
2	-0,9133	0,00270	-0,02649	-0,3287	-0,20267
3	-0,8951	-0,01129	-0,02044	-0,3214	-0,17772
4	-0,9299	0,02037	0,01800	-0,8883	-0,26265
5	-0,9362	0,02430	0,06539	-0,8995	-0,28255
6	-0,9240	0,02867	0,07868	-0,8882	-0,26591
7	-0,9262	0,02683	0,00677	-0,8948	-0,27936
8	-0,9344	0,04891	0,06100	-0,8981	-0,27451
9	-0,9203	0,03827	0,05974	-0,9045	-0,27820
10	-0,9367	0,02823	0,04644	-0,9076	-0,27883
Нова 1	-0,9327	0,01571	0,15592	-0,9072	-0,27496
Нова 2	-0,9316	0,02617	0,09776	-0,9078	-0,27583
Нова 3	-0,9281	0,04715	0,16311	-0,9086	-0,27435
Нова 4	-0,9229	0,04346	0,14530	-0,9095	-0,27747
Нова 5	-0,9211	0,03981	0,18024	-0,9131	-0,27920
Нова 6	-0,9326	0,03498	0,12592	-0,9163	-0,28652
Нова 7	-0,9298	0,05401	0,13759	-0,9231	-0,28774
Нова 8	-0,9184	0,02574	0,20173	-0,9204	-0,27198
Нова 9	-0,9200	0,06748	0,18150	-0,9230	-0,28410
Нова 10	-0,9235	0,09543	0,17181	-0,9256	-0,28923
Нова 11	-0,8994	0,10865	0,22821	-0,9270	-0,28758
Нова 12	-0,8508	0,08934	0,31835	-0,9261	-0,28418
Нова 13	-0,9194	0,10535	0,08860	-0,9284	-0,29927
Нова 14	-0,9212	0,10979	0,11372	-0,9272	-0,30349

802–1050 нм, другий фактор – корельований з початковими ознаками, що відповідають довжинам хвиль 718–753 нм і третій фактор – корельований з початковими ознаками, що відповідають довжинам хвиль 507–540 нм.

Для миші перший фактор найбільш корельований з початковими ознаками, що відповідають довжинам хвиль 414–504 нм, 547–683 нм і 712–850 нм; другий фактор – корельований з початковими ознаками, що відповідають довжинам хвиль 850–1050 нм; третій фактор – з початковими ознаками, що відповідають довжинам хвиль 510–535 нм; четвертий фактор – з початковими ознаками, що відповідають довжинам хвиль 437–480 нм; п'ятий фактор – з початковими ознаками, що відповідають довжинам хвиль 400–434 нм; шостий фактор – з початковими ознаками, що відповідають довжинам хвиль 353 нм,

370 нм і 976 нм; сьомий фактор – з початковими ознаками, що відповідають довжинам хвиль 350 нм і 996 нм; восьмий фактор – з початковими ознаками, що відповідають довжинам хвиль 350–360 нм.

Для амброзії перший фактор найбільш корельований з початковими ознаками, що відповідають довжинам хвиль 360–504 нм, 614–675 нм, 695–944 нм і 712–850 нм; другий фактор – корельований з початковими ознаками, що відповідають довжинам хвиль 495–510 нм, 535–593 нм, 678–690 нм, 947–1038 нм.

Для перевірки достатності розміру нового простору для кожної рослини, була проведена процедура розрахунку сумарної спільності і спільності по кожному фактору. Так, для амброзії сумарна спільність дорівнює 254,28, для першого фактору – 145,27, для другого фактору – 124,55. Тобто загальна спільність за двома факторами дорівнює $145,27 + 124,55 = 269,82$. Оскільки загальна спільність за двома факторами (269,82) більше ніж сумарна спільність (254,28), то розмірність ознакового простору у кількості двох обрано вірно. У табл. 7 наведено результат розрахунку спільності для амброзії. Графіки власних значень для кукурудзи, миші та амброзії наведені рис. 2–4.

Аналогічна процедура перевірки нових розмірностей була проведена для кукурудзи та миші. В результаті проведених розрахунків, розмірність ознакових просторів для кукурудзи та миші обрано вірно, оскільки для миші сума спільностей по факторам дорівнює 356,87, а сумарна спільність 224,78. Для кукурудзи: сума спільностей по факторам – 256,14, а сумарна спільність – 229,24.

6 ВИСНОВКИ

В роботі розглянута задача зменшення кількості інформаційних ознак за допомогою факторного аналізу, а саме методу головних компонентів.

В результаті проведення експериментальних досліджень встановлено, що можливо перейти до нового простору з меншою кількістю ознак при збереженні рівня інформативності не менше 80%. Так для кукурудзи

Таблиця 6 – Факторні навантаження для мишію

Змінні	Факторні навантаження (для мишію)							
	Фактор 1	Фактор 2	Фактор 3	Фактор 4	Фактор 5	Фактор 6	Фактор 7	Фактор 8
1	-0,0832	0,00065	-0,13211	0,259002	0,121068	0,141987	0,782968	-0,162445
2	-0,1140	-0,44176	-0,04021	-0,238139	-0,088848	-0,385649	-0,031048	0,477626
3	-0,0672	-0,41766	-0,03178	-0,320021	-0,026030	-0,079956	-0,299183	-0,486865
4	-0,2888	-0,64547	-0,03854	-0,235601	-0,035044	-0,041635	-0,060987	-0,088411
5	-0,3005	-0,60634	-0,03809	-0,231377	-0,056948	-0,041635	0,051629	0,100092
6	-0,3465	-0,60569	-0,10694	-0,226901	-0,054001	-0,167011	0,081305	0,191632
7	-0,3274	-0,65189	-0,05375	-0,200093	-0,124574	-0,062364	0,003541	-0,067721
8	-0,3494	-0,62809	-0,14527	-0,187012	-0,091447	-0,229901	-0,024393	-0,074242
9	-0,3689	-0,56764	-0,12693	-0,208454	-0,156561	-0,083712	0,064355	0,037291
10	-0,3340	-0,61098	-0,07223	-0,163582	-0,112814	-0,254564	0,072627	0,001780
Нова 1	-0,3455	-0,69667	-0,07004	-0,145086	-0,046614	0,036432	-0,012333	-0,037291
Нова 2	-0,3741	-0,67225	-0,14621	-0,171111	-0,099000	-0,089167	0,014638	0,001780
Нова 3	-0,4215	-0,69452	-0,18345	-0,149507	-0,070722	0,185336	-0,010266	0,020758
Нова 4	-0,4024	-0,66716	-0,13470	-0,159141	-0,091824	0,131594	-0,032505	-0,019124
Нова 5	-0,3559	-0,71232	-0,10534	-0,125823	-0,070673	0,001574	0,012386	0,117445
Нова 6	-0,4320	-0,67551	-0,11532	-0,144367	-0,127146	-0,024959	0,010995	-0,012405
Нова 7	-0,4828	-0,63595	-0,14191	-0,067666	-0,123461	0,020609	0,013794	-0,040740
Нова 8	-0,4398	-0,69172	-0,17675	-0,075642	-0,076265	0,158533	-0,037117	0,022666
Нова 9	-0,4870	-0,64593	-0,14645	-0,133559	-0,135250	0,132538	0,018388	0,037193
Нова 10	-0,5379	-0,62809	-0,12343	0,021363	-0,172884	0,028273	-0,012258	-0,042859
Нова 11	-0,5832	-0,62341	-0,17869	0,034503	-0,174038	0,156002	-0,036742	0,083318
Нова 12	-0,5296	-0,66663	-0,23535	0,018286	-0,136432	0,258610	-0,062037	0,079887
Нова 13	-0,6214	-0,53139	-0,06501	0,025591	-0,274416	0,009128	-0,000690	0,083146

Таблиця 7 – Результат розрахунку спільностей для амброзії

Змінна	Спільність для амброзії		
	Від фактора 1	Від фактора 2	Значення R-Square
1	0,345961	0,347358	0,951460
2	0,108021	0,149095	0,811392
3	0,103269	0,134854	0,855376
4	0,789116	0,858102	0,995631
5	0,809054	0,888888	0,997044
6	0,788982	0,859689	0,997014
7	0,800586	0,878628	0,997250
8	0,806656	0,882010	0,997462
9	0,818060	0,895454	0,998322
10	0,823716	0,901460	0,998506

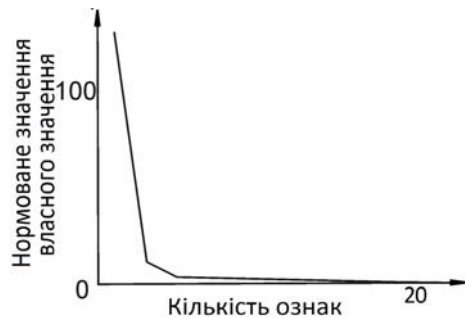


Рисунок 4 – Графік власних значень для амброзії

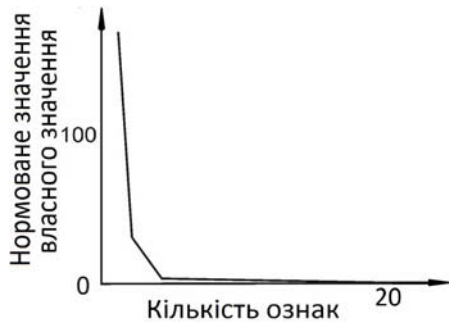


Рисунок 2 – Графік власних значень для кукурудзи

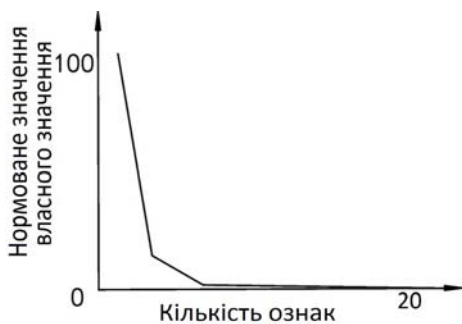


Рисунок 3 – Графік власних значень для мишію

розмірність нового простору дорівнює три, для мишію вісім, а для амброзії два. Різницю в кількості ознак для досліджуваних рослин, можна пояснити різною геометричною структурою листків, їхньою відмінністю у морфологічному та анатомічному складу, а це в свою чергу накладає відбиток на структурі багатоспектральних даних.

Практична цінність отриманих результатів полягає в тому, що результати розрахунків можуть бути корисними при побудові класифікатора рослинних об'єктів. Як зазначалось, при використанні факторного аналізу, окремі процедури виявлення ознак і класифікація не потрібні, вони вже закладені в самому методі., а тому класифікатор, в алгоритм розпізнавання рослин якого буде закладений метод факторного аналізу, буде мати більш просту схему побудови, а це в свою чергу дасть вигреш в економічному сенсі.

7 ПОДЯКИ

Стаття пов'язана з бюджетною науково-дослідною роботою Запорізького національного технічного університету «Методи, моделі та пристрої прийняття рішень в системах розпізнавання образів», номер державної реєстрації 0111U000059.

СПИСОК ЛІТЕРАТУРИ

1. Загоруйко Н.Г. Методы распознавания и их применение / Н. Г. Загоруйко. – М. : Сов. радио, 1972. – 270 с.
2. Дуда Р. Распознавание образов и анализ сцен / Р. Дуда, П. Харт. – М. : Мир, 1976. – 326 с.
3. Харман Г. Современный факторный анализ / Г. Харман. – М. : Статистика, 1972. – 488 с.
4. Бююль А. SPSS: Искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей / А. Бююль, П. Цефель. – СПб.: ООО «ДиаСофтЮП», 2002. – 603 с.
5. Факторный, дискриминантный и кластерный анализ: пер. с англ. / [Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др.; под ред. И. С. Енюкова. – М. : Финансы и статистика, 1989. – 215 с.
6. Прикладная статистика. Классификация и снижение размерности. Справочное издание / [С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин]. – М. : Финансы и статистика, 1989. – 607 с.
7. Купер К. Индивидуальные различия / К. Купер. – М. : Аспект Пресс, 2000. – 527 с.
8. Боровиков В. STATISTICA: искусство анализа данных на компьютере. Для профессионалов / В. Боровиков. – СПб: Питер, 2001. – 656 с.
9. Митина О. В. Факторный анализ для психологов / О. В. Митина, И. Б. Михайловская. – М. : Учебно-методический коллектор, Психология, 2001. – 169 с.
10. Краснополюсовский А. С. Факторный классификационный анализ за методом функционально-статистичних випробувань / А. С. Краснополюсовський // Радіоелектронні та комп'ютерні системи. – 2004. – № 4. – С. 46–50.

Стаття надійшла до редакції 01.09.2014.
Після доробки 27.01.2015.

Шама Е. А.

Аспирант кафедры радиотехники и телекоммуникации Запорожского национального технического университета, Запорожье, Украина

УМЕНЬШЕНИЕ РАЗМЕРНОСТИ ПРИЗНАКОВОГО ПРОСТРАНСТВА РАСТИТЕЛЬНЫХ ОБЪЕКТОВ С ПОМОЩЬЮ ФАКТОРНОГО АНАЛИЗА

В статье показана возможность уменьшения числа информативных признаков для многоспектральных данных растительного происхождения с помощью методов факторного анализа. Для исследования были использованы реальные коэффициенты отражения кукурузы, мышея и амброзии. Перед использованием метода факторного анализа была выполнена процедура проверки соответствия начального набора данных нормальному закону распределению. Факторный анализ проводился отдельно для трех растений: кукурузы, мышея и амброзии. Для каждого растения установлены размеры нового пространства, а также предоставлена качественная оценка корреляции между элементами нового и начального пространства данных. Переход к новому признаковому пространству производился при условии, что уровень информативности поддерживался на уровне не менее 80% по отношению к начальному набору данных. Для каждого вида растения была проведена процедура проверки правильности выбора количества признаков с помощью расчета общностей. Полученные результаты могут быть использованы при построении классификатора растительных объектов.

Ключевые слова: факторный анализ, признаковое пространство, спектральные коэффициенты яркости, признаки, классификатор.

Shama E. O.

Post-graduate Student of Department of Radio Engineering of Zaporizhzhya National Technical University, Ukraine

THE FEATURE SPACE DIMENSIONALITY REDUCTION OF PLANT OBJECTS BASED ON FACTOR ANALYSIS

A possibility of reducing the number of informative features for multispectral phytogenous data by means of the factor analysis method is shown in the article. For investigation there were used real reflection coefficients of maize, bristlegress and ambrosia. Prior to usage of the factor analysis it was performed the procedure to verify an initial set of data for conformity with distribution by the normal distribution law. The factor analysis was conducted separately for three plants: maize, bristlegress and ambrosia. Sizes of a new space were set for each plant, and also there was given quality estimation of correlation between elements of a new and initial space of the data. Transfer to new features was executed provided that a level of informativeness was maintained at a level of 80% minimum in comparison with the initial set of data. For each plant species the procedure of verification of the feature number choice correctness was performed by calculating of generalities. The obtained results can be used to construct plant objects classifier.

Keywords: factor analysis, feature space, spectral coefficients of brightness, feature, classifier.

REFERENCES

1. Zagoruyko N. G. Metody raspoznavaniya i ih primeneniye. Moscow, Sov. Radio, 1972, 270 p.
2. Duda R., Hart P. Raspoznavanie obrazov i analiz stsen. Moscow, Mir, 1976, 326 p.
3. Harman G. Sovremenniy faktorniy analiz. Moscow, Statistica. 1972, 488 p.
4. Byuyul A., Cefel P. SPSS: Iskustvo obrabotki informaci. Analiz statisticheskikh dannuh i vosstanovlenie skrutuh zakonornostey. SPb, ООО «DiaSoftYUP», 2002, 603 p.
5. Kim Dzh.-O., Muller Ch. U. Faktorniy analiz: statisticheskir metodu i prakticheskie, *Faktorniy, diskriminantniy i klasterniy analiz: sbornik rabot*; per. s angl.; Pod. red. I. S. Enyukova. Moscow, Finansu i sttistika, 1989, 215 p.
6. Aivazyan S. A., Buhstaber V. M., Yenyukov I. S., Meshalkin L. D. Prikladnaya statistika. Klassifikatsiya i snizheniye razmernosti. Spravochnoe izdaniye. Moscow, Finansy I statistika, 1989, 607 p.
7. Kuper K. Individualnie razlichiya. Moscow, Aspekt Press, 2000, 527 p.
8. Borovikov V. STATISTICA: iskusstvo analiza dannuh na komputere. Dlya professionalov, SPb: Piter, 2001, 656 p.
9. Mitina O. V., Mihaylovskaya I. B. Faktorniy analiz dlya psihologov. Moscow. Uchebno-metodicheskiy kolektor. Psihologiya, 2001, 169 p.
10. Krasnopoyasovskiy A. S. Faktorniy klasifikatsiynui analiz za metodom funnktsionalno-statustichnuh vuprobuvan, *Radioelektronni ta komputerni systemu*, 2004, No.4, pp. 46–50.