

УДК 528.06

С.І. АЛЬПЕРТ*

НОВІ ПІДХОДИ ДО ЗАСТОСУВАННЯ РІЗНИХ МЕТОДІВ ЗНАХОДЖЕННЯ БАЗОВИХ МАС ПРИ КЛАСИФІКУВАННІ ГІПЕРСПЕКТРАЛЬНИХ КОСМІЧНИХ ЗОБРАЖЕНЬ

*Науковий Центр аерокосмічних досліджень Землі ІГН НАН України, м. Київ, Україна

Анотація. На даний час визначення базових мас залишається важливою задачею, яка впливає на остаточні результати класифікування. Загального підходу для визначення базових мас не існує. У даній роботі були розглянуті різні методи визначення базових мас. У статті описується новий метод класифікування гіперспектральних космічних зображень, який відрізняється від відомих розв'язків однойменних задач особливим способом розбиття спектрального ознакового простору. Було проведено оптимізацію границь між класовими інтервалами на спектральних вісях за допомогою спеціального функціонала. Новий метод застосовує теорію свідчень Демпстера-Шейфера й визначає базові маси за допомогою частотного підходу. Також у даній статті був розглянутий новий підхід до знаходження базових мас з використанням відстані Махаланобіса. Була наведена формула, яка показує взаємозв'язок між відстанню Махаланобіса та базовими масами. Запропоновано метод визначення базових мас із використанням нечітких множин та нечітких чисел. Було зауважено, що теорія нечітких множин використовується при невизначених та неточних даних. Також описано метод визначення базових мас із використанням центральних вибірок. Зауважено, що для вибору центральних вибірок застосовуються опуклі оболонки. Потім визначаються базові маси з використанням відстані між класифікованими даними та вибраними центральними вибірками. Далі базові маси комбінуються за правилом Демпстера, і отримуються остаточні базові маси. Було наголошено на тому, що складні гіпотези теж розглядаються. Також проаналізовано новий підхід до визначення базових мас із використанням відстані між класифікованими даними та нормальним розподілом кожної характеристики для кожного еталонного класу. Описані методи отримання базових мас можуть бути застосовані при класифікуванні гіперспектральних космічних зображень, в управлінні транспортом, при класифікуванні рослинності, у клінічній діагностиці, при аналізі ризиків, розв'язанні тематичних завдань та при контролі якості води.

Ключові слова: базова маса, відстань Махаланобіса, теорія нечітких множин, центральна вибірка, класифікування зображень.

Аннотация. В настоящее время определение базовых масс остается важной задачей, которая влияет на окончательный результат классификации. Не существует общего подхода к определению базовых масс. В данной работе рассмотрены различные методы определения базовых масс. В статье описывается новый метод классификации гиперспектральных космических изображений, который отличается от решений одноименных задач особым способом разбиения спектрального признакового пространства. Было проведено оптимизацию границ между классовыми интервалами на спектральных осях с помощью специального функционала. Новый метод использует теорию свидетельств Демпстера-Шейфера и определяет базовые массы с помощью частотного метода. Также в статье рассматривается новый подход к нахождению базовых масс с помощью расстояния Махаланобиса. Была приведена формула, которая описывает связь между расстоянием Махаланобиса и базовыми массами. Предложен метод определения базовых масс с помощью нечетких множеств и нечетких чисел. Было отмечено, что теория нечетких множеств используется при наличии неопределенных и неточных данных. Также описан метод нахождения базовых масс с использованием центральных выборок. Было отмечено, что выпуклые оболочки используются для выбора центральных выборок. Потом базовые массы комбинируются по правилу Демпстера для того, чтобы получить окончательные базовые массы. Отмечено, что сложные гипотезы тоже рассматриваются. Также был проанализирован новый подход к определению базовых масс с использованием меры расстояния между классифицированными данными и нормальным распределением каждой характеристики для каждого эталонного класса. Описанные методы нахождения базовых масс могут быть использованы при классификации гиперспектральных космических изоб-

ражений, в управленні транспортом, при класифікації растительности, в клінічній діагностиці, при аналізі ризиків, вирішенні тематических задач и при контролі якості води.

Ключевые слова: базова маса, відстань Махаланобіса, теорія нечітких множин, центральна вибірка, класифікація зображень.

Abstract. This paper provides a description of a certain computer equipment components which allow remotely gain unauthorized access to computers. These components are Intel Management Engine (Intel ME) and Intel AMT. Intel ME is an independent subsystem which is included in almost all Intel processor chips since 2008. The chip is always connected to a power source, so the subsystem continues to work even when the computer is turned off. Vulnerabilities were discovered in Intel AMT, thereafter many computers using Intel processors became available for remote and local intruders. The paper also describes Chinese microchips that have been implemented into Supermicro equipment. This equipment was supplied not only to US commercial organizations but to governmental as well. Supermicro Chinese microchips have the ability to edit the code stream that heads to the processor by inserting their own code, or else it can change the instructions order for the processor. The microchip can intercept communication security coding, as well as prevent the restoration of the security system as a whole. The paper also provides an overview for recent sensational vulnerabilities of Meltdown, Spectre and ZombieLoad in Intel and ARM processors that allows manipulating a computer in a varying degree. These vulnerabilities are similar to each other, they allow a malicious application to read any type of computer memory, including kernel. It became feasible thanks to a speculative code execution system. Both personal user data can be stolen, such as browser history, website content, passwords, or system data, and disk encryption keys. Security experts should take into account the points above, as in certain cases all these things could possibly turn into national scale problems, both financial and political.

Keywords: vulnerability of computer systems, information security, Intel ME, Intel AMT, Supermicro microchip, vulnerability of Meltdown, Specter and ZombieLoad.

DOI: 10.34121/1028-9763-2020-1-30-42

1. Вступ

Мета даної роботи полягає в аналізі та порівнянні нових підходів до застосування різних методів знаходження базових мас при вирішенні задач класифікування супутникових зображень. Розглядається новий метод класифікування, що включає в себе оптимізацію границь між класами при розбитті ознакового простору за допомогою спеціального функціонала та знаходження базових мас за допомогою частотного підходу. Описується метод знаходження базових мас, який застосовується при вирішенні задач класифікування лісів та сільськогосподарських територій, з використанням метрики Махаланобіса. Проводиться аналіз методів знаходження базових мас із використанням теорії нечітких множин, оскільки при розв'язанні задач класифікування ми часто маємо оперувати з неповною та суперечливою інформацією [1]. Також розглядається метод знаходження базових мас із використанням опуклих оболонок, який має широке застосування у клінічній діагностиці, у моніторингу та аналізі навколишнього середовища, у сфері управління транспортом, в аналізі ризиків, у задачах класифікування рослинності та моніторингу якості води. Описується та аналізується метод знаходження базових мас із використанням міри відстані між класифікованими та еталонними даними. Даний метод застосовується в задачах класифікування рослинності.

2. Новий метод оптимізації границь між класами при розбитті ознакового простору та знаходження базових мас за допомогою частотного підходу

2.1. Новий алгоритм оптимізації границь

Спочатку нам треба кожен клас представити у вигляді деякої області (інтервалу) у просторі спектральних ознак. Для цього ми проводимо розбиття ознакового простору. При формуванні даних областей (інтервалів) використовуються значення сигналів пікселів навча-

льної вибірки. Розмірність області визначається числом зональних зображень, які використовуються для проведення процедури класифікування. Межі сформованих інтервалів визначаються величинами сигналів пікселів класу у спектральному каналі. Також слід зауважити, що побудовані інтервали не перетинаються, але межують один з одним. Число класів L визначає загальну кількість інтервалів у спектральному каналі.

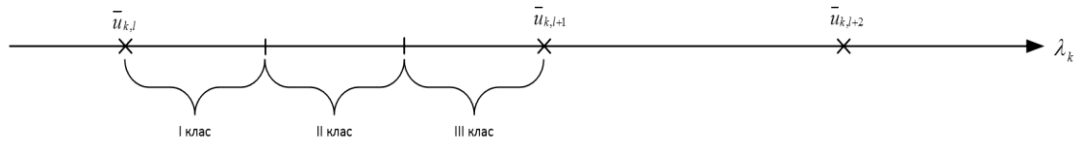


Рисунок 1 – До визначення положення границь, що розділяють класи на спектральній вісі

Введемо позначення: n_i – кількість елементів i -го класу, що знаходяться в інтервалі між середніми значеннями сигналів пікселів $\bar{u}_{k,l}$ та $\bar{u}_{k,l+1}$.

Відповідно, n_{i+1} – кількість елементів $i + 1$ -го класу. Наша задача полягає в максимізації функціоналу $\lg \frac{n_i}{n_{i+1} + 1}$ для інтервалу між середніми значеннями сигналів пікселів

$\bar{u}_{k,l}$ та $\bar{u}_{k,l+1}$ з метою знаходження нових оптимальних границь між $\bar{u}_{k,l}$ та $\bar{u}_{k,l+1}$. Нехай

n_1 – кількість елементів 1-го класу, n_2 – кількість елементів 2-го класу для інтервалу між $\bar{u}_{k,l}$ та $\bar{u}_{k,l+1}$. Розраховуємо функціонал $\lg \frac{n_1}{n_2 + 1}$. У даному інтервалі викидаємо крайній

елемент справа. Отримаємо нові значення n_1 та n_2 . Позначимо їх як n_1^* та n_2^* . Тоді

розраховуємо нове значення функціонала $\lg \frac{n_1^*}{n_2^* + 1}$. Рухаємося далі у напрямку справа на-

ліво в інтервалі між значеннями $\bar{u}_{k,l}$ та $\bar{u}_{k,l+1}$ і викидаємо наступний крайній об'єкт, який знаходиться найближче до правої границі (тобто, вже викидаємо два крайніх елементи справа). Знов розраховуємо нові значення кількості об'єктів 1-го та 2-го класів: n_1^{**} та n_2^{**}

і нове значення функціонала $\lg \frac{n_1^{**}}{n_2^{**} + 1}$. Продовжуємо дану процедуру, рухаючись справа

наліво в даному інтервалі, поки не дійдемо до останнього елемента. Потім порівнюємо всі знайдені на попередніх етапах значення функціонала та обираємо його максимальне значення, яке і буде одними із нових границь в інтервалі між значеннями $\bar{u}_{k,l}$ та $\bar{u}_{k,l+1}$.

Тепер у даному інтервалі рухаємося у протилежному напрямку, зліва направо, аналогічно викидаючи поступово елементи, розраховуючи нові значення кількості об'єктів та відповідно нові значення функціонала. Потім обираємо знов максимальне значення даного функціонала, яке і буде другою новою оптимальною границею для даного інтервалу. Тобто, інтервал між значеннями $\bar{u}_{k,l}$ та $\bar{u}_{k,l+1}$ розбито на субінтервали. Два максимальних значення функціонала визначають нові границі між середніми значеннями сигналів пікселів $\bar{u}_{k,l}$ та $\bar{u}_{k,l+1}$. Знайшовши дві нові границі, інтервал буде розбиватися на 3 субінтервали, незалежно від кількості класів. При цьому в першому субінтервалі будуть представники 1-го класу, у другому субінтервалі будуть представники всіх класів, у третьому субінтервалі знаходяться представники 3-го класу. Далі проводимо аналогічну процедуру розбиття на субінтервали для наступного інтервалу. Визначаємо базові маси за допомогою часто-

тного підходу, а саме формули (1) для усіх субінтервалів, крім середніх, оскільки для всіх середніх субінтервалів базова маса від складних гіпотез завжди буде рівна 1.

2.2. Частотний підхід обчислення базових мас для класів

Припустимо, що у j -му інтервалі (області) ми маємо M_j пікселів, при цьому M_1 – це число пікселів класу, для якого даний інтервал (область) є власним, а M_2 – загальне число пікселів інших класів. Оцінка базової ймовірності піксельних об'єктів у власному інтервалі вказаного класу визначається за такою формулою:

$$m^{(c)} = \frac{M_1}{M_j}, \quad (1)$$

сумарна базова ймовірність піксельних об'єктів інших класів у цьому ж інтервалі визначається за такою формулою:

$$m^{(nc)} = \frac{M_2}{M_j}. \quad (2)$$

Таку ж саму вищеописану процедуру проводимо з іншими спектральними каналами. Отримуємо нове розбиття ознакового простору з новими оптимізованими границями. Проводимо процедуру класифікування, використовуючи правило комбінації Демпстера та поняття пігністичної ймовірності [2–3].

2.3. Процедура класифікування

Припустимо, піксельний об'єкт π_n характеризується векторним сигналом $\mathbf{u}_n = (u_{1n}, u_{2n}, \dots, u_{kn}, \dots, u_{Kn})$ вже з урахуванням нового розбиття ознакового простору.

Крок 1. У кожному спектральному каналі з новими оптимізованими границями визначається субінтервал, якому відповідає певний сигнал u_{kn} пікселя π_n ($k = 1, 2, \dots, K$).

Крок 2. Визначаються імена усіх класів об'єктів, що знаходяться у кожному такому субінтервалі, після чого отримуються значення базової ймовірності об'єктів даного субінтервалу.

Крок 3. Розраховується за допомогою правила Демпстера комбіноване значення мас для всіх гіпотез.

Крок 4. Розраховуються пігністичні ймовірності VetP для всіх гіпотез.

Крок 5. Проранжовуємо отримані значення пігністичних ймовірностей даних гіпотез, далі визначаємо найбільш імовірний клас, до якого належить піксель, використовуючи критерій максимуму пігністичної ймовірності.

Преваги даного методу: метод забирає небагато часу та досить нескладний для обчислень. Недоліки: даний метод дає не досить точні результати при роботі з нечіткими множинами і не може оперувати суперечливими даними за наявності великого значення коефіцієнта конфліктності.

Псевдокод

S : number of spectral bands

K : number of classes and number of intervals

n_k : number of elements of k -class for interval between average values m_k and m_{k+1} n_{k+1} :

number of elements of $k+1$ -class for interval between average values m_k and m_{k+1}

For $p := 1$ to S (йдемо по всіх спектральних каналах)

For $r := 1$ to K (йдемо по всіх інтервалах у кожному спектральному каналі)

procedure delFirst;

```

var
  a : array [1..N] of integer ;
  i : integer ; { змінна I вводитьься як індекс масиву }
  res : real;
begin
  Readln('N = ', N);
  For i:=1 to (N-1) do
    a[i]:=a[i+1];
    a[N]:=0;
    res:=ln(n[i]/(n[i+1]+1))/ln(10);
    Writeln(res);
end;
procedure delLast;
var
  a : array [1..N] of integer ;
  i : integer ; { змінна I вводитьься як індекс масиву }
  res : real;
begin
  For i:=(N-1) to 1 do
    a[i+1]:=a[i];
    a[i]:=0;
    res:=ln(n[i]/(n[i+1]+1))/ln(10);
    Writeln(res);
end;
begin
  delFirst;
  delLast;
end.

```

L_1 : number of functionals F_i of interval between average values m_k and m_{k+1} for right-to-left direction

```

function max(mas:array of functionals  $F_i$ );
begin
   $F_i := mas[1]$ 
  for  $i:=1$  to  $L_1$  do
  begin
    if  $F_i < mas[i]$  then
       $F_i := mas[i]$ ;
  end;
  max :=  $F_i$ ;
end;

```

write (“we get a new first bound for interval between average values m_k and m_{k+1} for right-to-left direction”);
end.

L_2 : number of functionals F_i^* of interval between average values m_k and m_{k+1} for left-to-right direction

```

function max(mas:array of functionals  $F_i^*$ );

```

```

begin
   $F_i^* := mas[1]$ 
  for  $i := 1$  to  $L_2$  do
    begin
      if  $F_i^* < mas[i]$  then
         $F_i^* := mas[i]$ ;
      end;
       $max := F_i^*$ ;
    end;
    write (“we get a new second bound for interval between average values  $m_k$  and  $m_{k+1}$  for
left-to- right direction”);
    end.
    write (“we get new bounds for interval between average values  $m_k$  and  $m_{k+1}$ , so we divide
interval into  $K^*$  subintervals”);
    end.
    end.
    end.
     $K^*$  : number of subintervals
    For  $p := 1$  to  $S$  (йдемо по всіх спектральних каналах)
    For  $j := 1$  to  $K^*$  (йдемо по всіх субінтервалах у кожному інтервалі кожного спектра-
льного каналу)
    begin
      writeln (“рахуємо базові маси гіпотез:  $m_p(A_k), m_p(A_{k+1}), m_p(A_{k+2}), \dots$  для всіх субінте-
рвалів спектрального каналу, при цьому враховуємо, що в середніх субінтервалах кожного
інтервалу будуть представники всіх класів, а базова маса від цих всіх складних гіпотез
завжди буде рівна 1”);
      readln ( $m_p(A_k), m_p(A_{k+1}), m_p(A_{k+2}), \dots$ );
      end.
      end.
      end.
      For  $i := 1$  to  $S$  do (йдемо по всіх спектральних каналах)
      begin
        writeln (“знаходимо нові базові маси для всіх гіпотез, комбінуючи базові маси субі-
нтервалів даних спектральних каналів за правилом Демпстера:  $m(A_k), m(A_{k+1}), m(A_{k+2}), \dots$ 
”);
        readln ( $m(A_k), m(A_{k+1}), m(A_{k+2}), \dots$ );
        end.
        begin
          writeln (“шукаємо пігністичну ймовірність для всіх простих гіпотез:
 $BetP(A_k), BetP(A_{k+1}), BetP(A_{k+2}), \dots$ ”);
          readln ( $BetP(A_k), BetP(A_{k+1}), BetP(A_{k+2}), \dots$ );
          end.
        end.
      end.
    end.
  end.
end.

```

3. Метод знаходження базових мас за допомогою розрахунку відстані Махаланобіса до класів

Припустимо, що ми маємо багатоспектральний знімок G , який включає в себе K зональних зображень:

$$G = \{S_k\}; k = 1, 2, \dots, K. \quad (3)$$

Усі зональні зображення мають однакове число пікселів N .

k -те зональне зображення S_k можна представити у такому вигляді:

$$S_k = \{\pi_n, u_{nk}\}; n = 1, 2, \dots, N, \quad (4)$$

де u_{nk} – рівень сигналу n -го пікселя π_n зонального зображення S_k .

Зауважимо, що один і той же піксель на різних зональних зображеннях має різні рівні сигналу, тобто повний сигнал пікселя u_n – це K -вимірний вектор:

$$u_n = \{u_{nk}\}; k = 1, 2, \dots, K, \quad (5)$$

$u_n \in R^k$, компонентами якого є значення (рівні) цього сигналу на вісях спектрального простору R^k (спектральних вісях).

Даний метод полягає в тому, щоб провести попінкельну класифікацію багатоспектрального знімку G . Відома загальна кількість класів J . При цьому для кожного класу є достатня кількість репрезентів у навчальній вибірці Ω .

Спочатку відкидаємо малоінформативні спектральні канали методом оцінки кореляції і отримуємо K^* найбільш ефективних та інформативних спектральних каналів.

Далі розглянемо послідовність розв'язку даної задачі:

1. Беремо перше зональне зображення S_1 . Значення сигналів пікселів на цьому зображенні розташовуються на спектральній вісі r_1 . Використовуючи навчальну вибірку Ω , знайдемо середні значення сигналів пікселів для кожного із класів та побудуємо границі між класами на спектральній вісі, що розглядається. Границі між суміжними класами будемо, використовуючи метод оптимізації границь між класами при розбитті ознакового простору за допомогою максимізації функціонала $\lg \frac{n_i}{n_{i+1} + 1}$, що описаний у I-му пункті.

Таким чином, уздовж спектральної вісі було виділено не тільки два класи – перший та другий, але ще і додатковий змішаний клас, який був створений пікселями першого та другого класів. Аналогічним чином може з'явитися додатковий змішаний клас між другим та третім класами і т.д. Та ж сама процедура проводиться для усіх класів на першій спектральній вісі. Потім беремо другу спектральну вісь і проводимо аналогічну процедуру.

У результаті отримуємо у K^* -мірному спектральному просторі області із сигналів пікселів із навчальної вибірки: область сигналів пікселів першого класу, другого і т.д., ще і змішані області, що утворюються сигналами пікселів різних класів (як перетини класів). Кожну з таких областей можна розглядати як багатовимірний паралелепіпед, в якому знаходяться сигнали пікселів класу чи декількох класів.

2. Також слід зазначити, що сигнал у будь-якій одній області – це вектор з K^* складовими елементами. Кількість таких сигналів у довільній області може бути – Q . Ці сигнали утворюють множину випадкових сигналів, для якої можна обчислити коваріаційну матрицю. При цьому коваріаційні матриці розраховуються окремо для області з пікселів

кожного класу та для кожної змішаної області, яку утворюють пікселі з двох чи більше класів. Також для кожної області знаходимо центр тяжіння (середнє значення). Коваріаційна матриця для p -го класу буде квадратною. При цьому число рядків та число стовпчиків даної матриці буде дорівнювати кількості зональних зображень K^* .

$$S_p = \begin{bmatrix} \delta_1^2 & \rho_{12}\delta_1\delta_2 & \dots & \rho_{1K^*}\delta_1\delta_{K^*} \\ \rho_{21}\delta_2\delta_1 & \delta_2^2 & \dots & \rho_{2K^*}\delta_2\delta_{K^*} \\ \dots & \dots & \dots & \dots \\ \rho_{K^*1}\delta_{K^*}\delta_1 & \rho_{K^*2}\delta_{K^*}\delta_2 & \dots & \delta_{K^*}^2 \end{bmatrix}. \quad (6)$$

3. Тепер, якщо взяти будь-яку довільну точку A , яка описується вектором $\mathbf{u}^* = (u_1^*, u_2^*, \dots, u_{K^*}^*)$, то можемо розрахувати відстань між цією точкою та областю за допомогою метрики Махаланобіса:

$$MD_p = \sqrt{(u^* - \mu_p) \cdot S_p^{-1} \cdot (u^* - \mu_p)^T}, \quad (7)$$

де μ_p – математичне сподівання p -го класу ($p = 1, 2, \dots, J$), J – загальна кількість класів, S_p^{-1} – обернена коваріаційна матриця для p -го класу.

4. Використовуючи знайдену метрику Махаланобіса, знаходимо базові ймовірності (маси). Враховуючи, що відстань Махаланобіса приймає значення від “0” до нескінченності, а значення ймовірностей знаходяться в діапазоні від “0” до “1”, застосовуємо таке перетворення. Нехай у нас є деякий піксель, для якого треба визначити базову ймовірнісну масу. Припустимо, що даний піксель представлено у спектральному просторі деякою точкою A . Нам треба розрахувати відстань MD_j^A між цією точкою та центром тяжіння області ω_j класу. Для цього спочатку обчислюємо відстань Махаланобіса між цією точкою та кожною з областей-класів. Далі знаходимо середнє значення з отриманих відстаней Махаланобіса – \overline{MD} . Потім відстань MD_j^A між точкою A та областю ω_j класу перетворюємо у базову ймовірність точки A за такою формулою:

$$m(A) = 1 - \frac{1}{1 + e^{-k(MD_j^A - \overline{MD})}}, \quad (8)$$

де k – коефіцієнт.

Після нормалізації розраховуються пігністичні ймовірності для кожної із областей-класів. Потім пікселю, який ми маємо класифікувати, надається той клас, для якого пігністична ймовірність буде найбільшою. Для даного пікселя задача класифікування розв’язана. Далі переходимо до наступного пікселя і проводимо таку ж саму процедуру. Переваги даного методу: метод нескладний для обчислень і забирає небагато часу. Недоліки: даний метод не підходить при роботі з нечіткими множинами при наявності невизначеної, неповної та неточної інформації [2–4].

4. Метод знаходження базових мас із використанням теорії нечітких множин

Досить часто при розв’язанні задач класифікування ми маємо справу з неповною, неточною та суперечливою інформацією. Теорія нечітких множин дає змогу описувати, обробляти та встановлювати взаємозв’язки між неточними даними. Наведемо основні поняття та означення з теорії нечітких множин [5]. Нехай X – універсальна множина. Тоді нечітка підмножина $A \subset X$ визначається за допомогою характеристичної функції

$$\mu_A : X \rightarrow [0, 1]. \quad (9)$$

При цьому кожному елементу $x \in A$ надається дійсне значення характеристичної функції $\mu_A(x)$ із інтервалу $[0, 1]$. Значення $\mu_A(x)$ вказує на вагу (grade of membership) елемента x у множині A . Нехай маємо нечітку множину $A \subset X$ та будь-яке дійсне число $\alpha \in [0, 1]$. Тоді α – рівень або α – зріз, що відповідає множині A , позначається ${}^\alpha A$ і є чіткою множиною:

$${}^\alpha A = \{x \in X : \mu_A(x) \geq \alpha\}. \quad (10)$$

Трикутне нечітке число A визначається через трійку чисел $[a, b, c]$. Його характеристична функція має вигляд

$$\mu_A(x) = \begin{cases} \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \end{cases}. \quad (11)$$

Центральний елемент описується за допомогою α – зрізу таким чином:

$${}^\alpha A_i = \{x : \mu_A(x) \geq \alpha_i\} = [{}^\alpha A_{i, \text{lower}}, {}^\alpha A_{i, \text{upper}}], \quad (i = 1, 2, 3, \dots, n), \quad (12)$$

де $\alpha_0 = 1 > \alpha_1 > \alpha_2 > \dots > \alpha_n = 0$, $\alpha \in [0, 1]$.

Перший підхід до знаходження базових мас

У даному випадку ми беремо інтеграл від характеристичної функції між двома різними α – зрізами, отримуємо базову масу ${}^\alpha A_i$ (α – зрізу):

$$m({}^\alpha A_i) = \frac{\int_{{}^\alpha A_{i, \text{lower}}}^{{}^\alpha A_{i, \text{upper}}} \mu_{(A)}(x) dx}{\sum_i \int_{{}^\alpha A_{i, \text{lower}}}^{{}^\alpha A_{i, \text{upper}}} \mu_{(A)}(x) dx}. \quad (13)$$

Другий підхід до знаходження базових мас

Базова маса визначається за формулою

$$m({}^\alpha A_i) = \frac{\alpha_{i-1}(1 - \alpha_i)}{\sum_i \alpha_{i-1}(1 - \alpha_i)}. \quad (14)$$

Третій підхід до знаходження базових мас

У третьому методі α – зріз розглядається як інтервал довіри з рівнем довіри: $1 - \alpha$. Базова маса визначається за формулою

$$m({}^\alpha A_i) = \frac{1 - \alpha_i}{\sum_i (1 - \alpha_i)}. \quad (15)$$

Наведені три методи знаходження базових мас мають такі переваги: використовуються при розв'язанні задач класифікування за наявності невизначеної, неточної та неповної інформації.

Основний недолік даних методів: наведена процедура знаходження базових мас не завжди є обґрунтованою, оскільки нам відомий тільки один параметр – трикутне нечітке число, яке вказує на максимальне, мінімальне та найбільш вірогідне значення й визначається через характеристичну функцію.

5. Метод визначення базових мас, заснований на мірі відстані між класифікованими даними та еталонними даними

Даний метод базується на використанні міри відстані між класифікованими та еталонними даними, тобто на відстані між класифікованими об'єктами та нормальним розподілом кожної характеристики для кожного еталонного класу. Цей метод використовує класифіковані дані для побудови нормального розподілу еталонних даних для кожної характеристики. Базові маси визначаються на основі міри відстані між класифікованими даними та моделлю для кожної характеристики для кожного еталонного класу [6]. Нехай ми маємо n класів та N об'єктів. При цьому кожен клас містить l об'єктів. Також для кожного класу існує певне число характеристик – k . Припустимо, що ми з першого класу випадковим чином обираємо m об'єктів та знаходимо середнє значення μ й середнє квадратичне відхилення δ для першої характеристики для першого класу. Аналогічним чином з інших класів обираємо m об'єктів та знаходимо середнє значення й середнє квадратичне відхилення також для першої характеристики. Потім будуємо графік нормального розподілу для першої характеристики для кожного класу. Аналогічно будуємо моделі нормального розподілу для всіх інших характеристик для всіх класів. Класифікування проводиться таким чином. Обираємо випадковим чином об'єкт, який треба класифікувати, з певними характеристиками та аналізуємо взаємозв'язок між цим об'єктом та нормальним розподілом кожної характеристики кожного класу. Для цього спочатку знаходимо міру відстані між класифікованими даними x_i та нормальним розподілом характеристики класу A :

$$d(A) = 2 \int_{\mu}^{x_i} p[x|\mu] dx, \quad (16)$$

де $p[x|\mu]$ – функція густини нормального розподілу характеристики класу A , яка обчислюється таким чином:

$$p[x|\mu] = \frac{1}{\sqrt{2\pi}\delta} \exp\left\{-\frac{1}{2}\left[\frac{x-\mu}{\delta}\right]^2\right\}, \quad (17)$$

де μ – середнє значення нормального розподілу характеристики класу A , δ – середнє квадратичне (стандартне) відхилення.

Функція похибки обчислюється таким чином:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (18)$$

Тепер обчислюємо міру відстані $d(A)$ як функцію похибки:

$$d(A) = erf\left[\frac{x_i - \mu}{\sqrt{2}\delta}\right]. \quad (19)$$

Отримане значення міри відстані $d(A)$ виражає ступінь відхилення між класифікованими даними x_i та моделлю нормального розподілу певної характеристики певного класу A .

Також слід зазначити, що чим більше значення міри відстані, тим менше ймовірність, з якою класифіковані об'єкти x_i належать класу A . Маємо $p(A) = 1 - d(A)$ – ймовірність належності об'єктів x_i класу A . Значення міри відстані $d(A)$ виражає міру відстані між певною характеристикою класифікованого об'єкта та нормальним розподілом певної характеристики у кожного класу. Після обчислення міри відстані $d(A)$ обчислюємо ймовірності належності класифікованих об'єктів різним класам за певною характеристикою. Нормуємо отримані ймовірності та отримуємо базові маси. Аналогічно знаходимо базові маси належності об'єктів різним класам за іншими характеристиками. Використовуючи правило комбінації Демпстера, отримуємо результуючі базові маси належності об'єктів класам.

6. Метод визначення базових мас із використанням центральних вибірок

1) Даний метод використовує навчальні дані для отримання центральних вибірок, які дають необхідну інформацію для гіпотез [6–7]. Центральні вибірки, у свою чергу, використовуються для знаходження базових мас. Центральна вибірка – це множина даних, які, відповідно до характеристик, належать певному класу. Якщо у нас m класів, то буде m центральних вибірок. Центральні вибірки, які відповідають простим та складним гіпотезам, отримуємо використовуючи метод k -середніх. При цьому центральні вибірки, які містять об'єкти одного класу, використовуються всі, а із центральних вибірок, які містять об'єкти різних класів, обираються тільки репрезентативні в залежності від значення оптимального відношення, яке отримуємо, використовуючи опуклі оболонки. Опукла оболонка – це множина даних, що належить відомому класу. Або опукла оболонка-довжина опуклого інтервалу, що містить множину точок, які належать певному класу (для одновимірного простору). У двовимірному та тривимірному просторах форма опуклих оболонок вимірюється за допомогою площі та об'єму відповідно.

2) Отримуємо опуклі оболонки, що відповідають простим гіпотезам і містять об'єкти одного класу: $CH(C_m)$, $m \in M$. Отримуємо також опуклі оболонки, що відповідають складним гіпотезам. Нам треба вибрати репрезентативні вибірки, які відповідають складним гіпотезам: $C_{N \subseteq M}$, $card(N) \geq 2$, використовуючи опуклі оболонки та оптимальне відношення. У цьому випадку ми розглядаємо подібність між простими опуклими оболонками $CH(C_m)$, $m \in N$, що входять у складні. Наприклад, центральна вибірка C_{ab} , $a, b \in M$ ($a \neq b, a \neq \emptyset, b \neq \emptyset$), що відповідає складній гіпотезі, містить об'єкти, які належать класам ω_a та ω_b . У цьому випадку оцінюємо оптимальне відношення R_{ab} . Якщо опуклі оболонки $CH(C_a)$ та $CH(C_b)$ перетинаються, то це означає, що серед даних у навчальній вибірці переважає більше подібність, ніж протиріччя. При цьому оптимальне відношення $R_{ab} = 1$. Якщо опуклі оболонки $CH(C_a)$ та $CH(C_b)$ не перетинаються, то у цьому випадку оптимальне відношення знаходиться за формулою

$$R_{ab} = \frac{CH(C_a) + CH(C_b)}{CH(C_a + C_b)}, \quad (20)$$

де $CH(C_a)$ – опукла оболонка, що містить точки із класу ω_a , $CH(C_b)$ – опукла оболонка, що містить точки з класу ω_b , тобто $CH(C_a)$ та $CH(C_b)$ – це довжини інтервалів, до

яких належать точки з класів ω_a та ω_b відповідно, $CH(C_a + C_b)$ – самий короткий неперервний інтервал, що містить $CH(C_a)$ та $CH(C_b)$. Оптимальне відношення R_{ab} виражає подібність між опуклими оболонками. Для складних гіпотез, коли $N \subseteq M$, $card(N) \geq 2$, встановлюємо порогове значення R_T . Якщо оптимальне відношення $R_N > R_T$, то розглядаємо дану складну гіпотезу та відповідну їй центральну вибірку. В іншому випадку, якщо $R_N < R_T$, то складні гіпотези та відповідні їм центральні вибірки, чие оптимальне відношення менше встановленого порогового значення, відкидаємо.

3) Обчислюємо відстань між класифікованими об'єктами та вибраними центральними вибірками.

4) Обчислюємо базові маси за формулами

$$m^s(\{C_{N_i}\}) = \alpha_{N_i} \varphi(d_i^s), \quad (21)$$

$$m^s(\{\Omega\}) = 1 - \sum_{i=1}^n \alpha_{N_i} \varphi(d_i^s), \quad (22)$$

де $0 < \alpha_{N_i} < 1$ – константа, яка надає вагу монотонно спадаючій функції φ , для якої виконується: $\varphi(0) = 1$ та

$$\lim_{d_i^s \rightarrow \infty} \varphi(d_i^s) = 0. \quad (23)$$

В експоненціальній формі:

$$\varphi(d_i^s) = \exp(-\gamma_{N_i} (d_i^s)^2), \quad (24)$$

де γ_{N_i} – константа, що відповідає центральній вибірці C_{N_i} .

5) Комбінуємо базові маси за правилом Демпстера.

6) Обчислюємо пігністичні ймовірності та визначаємо належність об'єкта до певного класу.

Переваги даного методу: базові маси обчислюються для всіх характеристик; є можливість уникнути сильних протиріч в отриманих базових масах, так як імовірності надаються як і простим, так і складним гіпотезам; базова маса надається множині Ω , завдяки чому тільки достовірні свідчення приймаються за прості гіпотези; оптимальне відношення R_N показує подібність між простими гіпотезами, що входять у складні. Чим вище обране нами порогове значення R_T , тим більше ми відкидаємо вибірки, що відповідають складним гіпотезам із високим ступенем внутрішнього протиріччя; даний метод дає можливість зменшити невизначеність за допомогою вибору репрезентативних центральних вибірок; завдяки тому, що властивості опуклості зберігаються не тільки в одновимірному, а також і в багатовимірному випадку, даний метод можна застосовувати до багатовимірних моделей, що описують характеристики об'єкта; для отримання точних результатів даний метод не потребує наявності великої кількості навчальних даних.

7. Висновки

У даній роботі було розглянуто нові підходи до застосування різних методів знаходження базових мас. Розглянуто метод класифікування, який відрізняється від відомих методів вирішення однойменних задач особливим способом розбиття спектрального ознакового простору з метою представлення класів областями у просторі спектральних ознак. При цьому показано, що числове значення ознаки класу у кожному спектральному каналі розташову-

ється у деякому інтервалі, границі якого визначаються за допомогою спеціального функціонала. Також було описано метод знаходження базових мас за допомогою метрики Махаланобіса. Для будь-якої точки за допомогою метрики Махаланобіса обчислюються відстані до кожної з областей, а потім ці відстані перераховуються у базові ймовірності. У статті було проведено аналіз методу знаходження базових мас із використанням теорії нечітких множин, оскільки при розв'язанні задач класифікування ми часто маємо оперувати з неповними, неточними та суперечливими даними. Розглядався метод знаходження базових мас із використанням опуклих оболонок та метод знаходження базових мас із використанням міри відстані між класифікованими та еталонними даними. Запропоновані методи знаходження базових мас можуть бути використані при вирішенні актуальних природоресурсних задач, таких як класифікування рослинності, пошук корисних копалин, класифікування лісів, сільськогосподарських земель, у медичній діагностиці, у сфері управління транспортом, контролю якості води та моніторингу навколишнього середовища [7].

СПИСОК ДЖЕРЕЛ

1. Beynon M.J., Curry B., Morgan P. The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modeling. *Omega: Int. Journal of Management Science*. 2000. Vol. 28. N 1. P. 37–50.
2. Smets P., Henrion M., Shachter R.D., Kanal L.N., Lemmer J. F. Constructing the pignistic probability function in a context of uncertainty. *Uncertainty in Artificial Intelligence*. 1990. Vol. 5. P. 29–40.
3. Shafer G. *A Mathematical Theory of Evidence*. Princeton: Princeton University Press, 1976. 297 p.
4. Альперт С.І. Порівняння нового удосконаленого підходу комбінування суперечливих даних з правилом Ягера. *Український журнал дистанційного зондування Землі*. 2018. № 17. С. 14–17. URL: <https://ujrs.org.ua/ujrs>.
5. Dutta P., Boruah H., Ali T. Fuzzy arithmetic with and without using α -cut method: a comparative study. *International Journal of Latest trends in Computing*. 2011. Vol. 2, N 1. P. 99–108.
6. Jiang W., Duanmu D., Fan X., Deng Y. A new method to determine basic probability assignment under fuzzy environment. *International Conference on Systems and Informatics (ICSAI 2012)*. P. 758–762.
7. Zhang C., Hu Y., Chan F.T.S., Sadiq R., Deng Y. A new method to determine basic probability assignment using core samples. *Knowledge-based Systems*. 2014. Vol. 69. P. 140–149.

Стаття надійшла до редакції 19.09.2019