

**УДОСКОНАЛЕНИЙ МЕТОД КОМБІНУВАННЯ ДАНИХ НА ОСНОВІ ТЕОРІЇ ДЕМПСТЕРА-ШЕЙФЕРА ЗА НАЯВНОСТІ СУПЕРЕЧЛИВИХ ДАНИХ**

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

---

***Анотація.** Правило комбінації Демпстера є найбільш відомим правилом комбінації, оскільки воно оперує з неповними та неточними даними. Але дане правило не може працювати за наявності занадто суперечливої інформації. Правило комбінації Мерфі може оперувати з суперечливими частинами свідчення, воно розраховує середню базову масу. Але цей метод також має певний недолік, оскільки не враховує взаємозв'язок між свідченнями. Тому було запропоноване нове вдосконалене правило комбінації, засноване на теорії Демпстера-Шейфера, яке просте для обчислень і має високу точність.*

***Ключові слова:** теорія свідчень, правила комбінації, суперечливі свідчення, правило Демпстера.*

***Аннотация.** Правило комбинации Демпстера является наиболее известным правилом комбинации, поскольку оно оперирует с неполными и неточными данными. Но данное правило не может работать при наличии слишком противоречивой информации. Правило комбинации Мерфи может оперировать с противоречивыми частями свидетельства, оно считает среднюю базовую массу. Но этот метод тоже имеет определенный недостаток, поскольку не учитывает взаимосвязь между свидетельствами. Поэтому было предложено новое усовершенствованное правило комбинации, основанное на теории Демпстера-Шейфера, которое простое для вычислений и имеет высокую точность.*

***Ключевые слова:** теория свидетельств, правила комбинации, противоречивые свидетельства, правило Демпстера.*

***Abstract.** Dempster's rule of combination is the most popular rule of combinations, because it processes incomplete and uncertain data. But this rule can not deal with highly conflicted information. Combination method of Murphy can deal with conflicting bodies of evidence, it calculates the mean of the basic probability assignment. But this method also has some disadvantage. Correlation among various evidences is not taken into account. It was proposed a new improved combination method. Based on the Dempster-Shafer theory, a new method has the advantage of simple calculation and high accuracy.*

***Keywords:** evidence theory, combination rules, conflicting evidence, Dempster's rule.*

## **1. Вступ**

За останнє десятиліття задача об'єднання даних, отриманих із різних джерел, стала особливо актуальною, оскільки знаходить широке застосування у таких сферах, як обробка, аналіз та класифікування зображень, вирішення сільськогосподарських, ресурсознавчих, екологічних задач, військова справа.

У даній статті буде розглянуто декілька методів об'єднання даних, отриманих із різних суперечливих джерел інформації. Буде наголошено на тому, що, не зважаючи на численні переваги правила комбінації Демпстера, його не можна застосовувати за наявності суперечливої інформації, оскільки у цьому випадку дане правило дає невірні результати.

Буде розглянуто метод усереднення Мерфі для підрахунку середньої базової маси гіпотез, який може оперувати із суперечливими даними. Описуються переваги та недоліки даного методу.

У даній роботі буде запропоновано новий удосконалений метод комбінування даних за наявності суперечливих джерел інформації, який, на відміну від методу Мерфі, враховує кореляцію між різними свідченнями та ефективно вирішує задачу комбінування даних [1–4].

## 2. Основні положення теорії Демпстера-Шейфера

За останній час застосування теорії Демпстера-Шейфера (ТДШ) дозволило досить успішно вирішити багато задач, які потребували об'єднання інформації із різних джерел.

Математичний апарат теорії Демпстера-Шейфера дає можливість працювати в умовах забрудненості та неповноти даних зі складними гіпотезами, пропонує просте правило для комбінування даних від різних джерел (спектральних каналів), дозволяє будувати інтервальні оцінки достовірності гіпотез. У ТДШ ключовим поняттям є поняття “маси” (mass), яке є узагальненням класичного поняття ймовірності. Масою можна відокремити поняття відсутності довіри від недовіри. Нехай  $A_0$  – обмежена множина, а  $A_i$  ( $i = 1, 2, \dots$ ) – його підмножини, тоді базова ймовірність визначається через функцію  $m$ :

$$\begin{cases} m(\emptyset) = 0, \\ \sum_{A_i \subseteq A_0} m(A_i) = 1, \quad (i = 0, 1, 2, \dots). \end{cases} \quad (1)$$

Нехай  $m_1$  і  $m_2$  – базові ймовірності гіпотези, отриманої з незалежних доведень, а  $A_{1i}$  та  $A_{2j}$  ( $i, j = 0, 1, 2, \dots$ ) – відповідні центральні елементи ( $m(A_{1i}) > 0, m(A_{2j}) > 0$ ). Тоді правило комбінації Демпстера задає нову базову ймовірність, яку можна представити за допомогою такої формули:

$$m(A_k) = \frac{\sum_{A_{1i} \cap A_{2j} = A_k} m_1(A_{1i})m_2(A_{2j})}{1 - C}, \quad (2)$$

де

$$C = \sum_{A_{1i} \cap A_{2j} = \emptyset} m_1(A_{1i})m_2(A_{2j}) \text{ – коефіцієнт конфліктності.} \quad (3)$$

Коефіцієнт конфліктності  $C$  вказує, наскільки суперечливими між собою є джерела інформації (спектральні канали).  $C$  лежить в інтервалі  $[0; 1]$ , чим сильніше ці протиріччя, тим ближче до одиниці стає величина  $C$  [5–11].

## 3. Комбінація даних, отриманих із суперечливих джерел інформації

Правило комбінації Демпстера має такі дві важливі математичні властивості, як комутативність та асоціативність.

Але правило комбінації Демпстера має певний недолік. Його не можна застосовувати при занадто суперечливих джерелах інформації, оскільки за наявності великого значення коефіцієнта конфліктності дане правило дає невірні результати.

Тому в 2000 році Мерфі було запропоновано метод для підрахунку середньої базової маси для кожної гіпотези. Цей метод усереднює усі базові маси, отримані з  $n$  частин свідчення, а потім ітераційно комбінує нові усереднені базові маси для кожної гіпотези за правилом комбінації Демпстера. У порівнянні з іншими підходами метод Мерфі може оперувати із суперечливими свідченнями.

Розглянемо випадок, коли в нас є  $n$  базові маси:  $m_1(a), m_2(a), \dots, m_n(a)$ .

Використовуючи підхід усереднення Мерфі, маємо:

$$m(a) = \frac{m_1(a) + m_2(a) + \dots + m_n(a)}{n}, \quad (4)$$

де  $a$  – деяка гіпотеза. Метод Мерфі працює з суперечливими джерелами інформації, але цей метод також має певний недолік. Він не враховує кореляцію (взаємозв'язок) між різ-

ними свідченнями, що, у свою чергу, призводить до незадовільного вирішення задачі об'єднання суперечливих даних. Метод усереднення Мерфі ефективно вирішує задачу комбінування суперечливих свідчень, але звичайне усереднення не завжди дає надійні результати, особливо у випадку, коли кількість свідчень є недостатньою для прийняття рішення. Причина цього полягає в тому, що просте усереднення не враховує асоціативний зв'язок між свідченнями, отриманими з різних джерел.

Тому у даному випадку пропонується використовувати новий модифікований метод комбінування даних за наявності великого значення коефіцієнта конфліктності на основі ТДШ [12–17].

#### 4. Новий модифікований метод комбінування даних за наявності великого значення коефіцієнта конфліктності

Представлений метод складається з 7 етапів.

1) Визначаємо відстань між двома частинами свідчення  $m_i$  and  $m_j$  :

$$d_{BPA}(m_i, m_j) = \sqrt{\frac{1}{2}(\|m_i\|^2 + \|m_j\|^2 - 2\langle m_i, m_j \rangle)}, \quad d_{BPA} \in [0, 1]. \quad (5)$$

$$\langle m_i, m_j \rangle = \sum_{k=1}^{2^n} \sum_{l=1}^{2^n} m_i(A_k) m_j(A_l) \frac{|A_k \cap A_l|}{|A_k \cup A_l|}, \quad k, l = 1, \dots, 2^n, \quad (6)$$

де  $\frac{|A_k \cap A_l|}{|A_k \cup A_l|}$  – коефіцієнт Жаккара двох множин  $A_k$  та  $A_l$ .

Коефіцієнт Жаккара дорівнює відношенню кількості елементів перетину множин до кількості елементів їхнього об'єднання. Був запропонований Полем Жаккаром у 1901 році.

У формулі (6) на коефіцієнт Жаккара множиться кожен доданок окремо.

$\|m_i\|^2 = \langle m_i, m_i \rangle$  – звичайна евклідова норма.

2) Визначаємо міру подібності між двома частинами свідчення  $m_i$  and  $m_j$  :

$$\text{sim}(m_i, m_j) = 1 - d(m_i, m_j). \quad (7)$$

3) Припускаємо, що число частин свідчення –  $n$ . Будуємо матрицю подібності:

$$SMM = \begin{bmatrix} 1 & s_{12} & \dots & s_{1n} \\ s_{21} & 1 & \dots & s_{2n} \\ \dots & \dots & \dots & \dots \\ s_{n1} & s_{n2} & \dots & 1 \end{bmatrix}. \quad (8)$$

4) Функція підтримки кожної частини свідчення  $m_i$  визначається, як

$$\text{Sup}(m_i) = \sum_{j=1}^n s_{ij}, \quad i = 1, 2, \dots, n. \quad (9)$$

5) Здійснюючи об'єднання даних, ми прагнемо дізнатися, яке саме свідчення (спектральний канал), що ми використовуємо, є надійним. Можливим є випадок, що одна частина свідчення є важливішою за іншу, і кожна частина свідчення має свою певну вагу у цілому свідченні. Наприклад, під час тривалих вимірювань деякі сенсори можуть бути більш надійнішими за інші завдяки високій стабільності. Таким чином, сенсори можуть відрізнятися за ступенем надійності. Ось чому процедура надання ваги свідченню є досить важли-

вою, особливо, коли свідчення, отримані з різних сенсорів, мають високий коефіцієнт конфліктності. У цьому випадку задача полягає в тому, щоб визначити серед усіх свідчень, яке з них є більш надійним, а якому можна приділити менше уваги. Надійним шляхом вирішення даної задачі є такий: якщо певна частина свідчення знаходить підтримку в інших свідченнях, то тоді дане свідчення буде мати більшу вагу, ніж та частина свідчення, яка має високий ступінь розбіжності (протиріччя) з іншими свідченнями. Беручи за основу дане правило, опишемо «ступінь надійності» свідчення, яке є вагою й відображає важливість кожної частини свідчення.

Надійність кожної частини свідчення  $m_i$  визначається, як

$$Crd(m_i) = \frac{\sup(m_i)}{\sum_{i=1}^n \sup(m_i)}. \quad (10)$$

Звідси також маємо  $\sum_{i=1}^n Crd_i = 1$ . Згідно з формулами (7) та (9), ми можемо зробити

висновок, що, чим коротша відстань від даної частини свідчення до інших частин свідчення, тим більшим є ступінь підтримки (функція підтримки) даної частини свідчення. Якщо частина свідчення отримує значну підтримку у інших частин свідчення, то її ступінь надійності є високим, і вона буде мати більший вплив на кінцевий результат об'єднання даних. Навпроти, якщо певна частина свідчення суперечить іншим свідченням, то її ступінь надійності – низький, отже, ця частина свідчення має значно менший вплив на кінцевий результат комбінації даних, отриманих з різних джерел. Таким чином, ступінь надійності – це вага, яка є індикатором важливості даного свідчення.

б) Коефіцієнт ваги  $m_M$  свідчення для певної гіпотези  $a$ :

$$m_M(a) = \sum_{i=1}^n Crd(m_i) \times m_i(a). \quad (11)$$

7) Далі, використовуючи правило Демпстера, комбінуюємо коефіцієнти ваги (11)  $n-1$  раз за умови, якщо ми маємо  $n$  частин свідчення.

Псевдокод

$n$ : number of images  $S_i$

For  $i:=1$  to  $n$

For  $j:=1$  to  $n$

$S$  is a vector of images  $S_i$

begin

write (' $S_i =$ ');

readln ( $S_i$ );

write (' $S_j =$ ');

readln ( $S_j$ );

writeln (“шукаємо відстань між спектральними каналами  $S_i$  та  $S_j$ :  $d_{BPA}(m_i, m_j)$ ”);

readln ( $d_{BPA}(m_i, m_j)$ );

end.

begin

writeln (“шукаємо функцію подібності між парами спектральних каналів  $S_i$  та  $S_j$ :

$sim(m_i, m_j)$ ”);

```

readln ( sim( $m_i, m_j$ ));
end.
Program Matrix (“будуємо матрицю подібності”)
type Tindex=1,...,  $n$ *
Vector=array [Tindex]of real;
Matrix =array[Tindex] of vector;
var sim: matrix;
action: real;
 $n, m$  ( $n = m$ ): tindex;
procedure Matrix Input  $S$  (var sim: Matrix);
var  $i, j$ : real;
begin
writeln (“Введіть”,  $n \times n$ , “елементів”);
for  $i:=1$  to  $n$  do
for  $j:=1$  to  $n$  do
begin
write ('sim[' $i, ', ', j, ']=$ );
read (sim[ $i, j$ ]);
end.
For  $i:=1$  to  $n$  do
begin
 $s:=0$ 
for  $j:=1$  to  $n$  do
 $s:=s+sim[i, j]$ 
writeln (“рахуємо функцію підтримки  $i$ -го свідчення, тобто рахуємо суму елементів у рядку матриці  $S$ ”);
end.
For  $i:=1$  to  $n$ 
 $S$  is a vector of images  $S_i$ 
Sort  $S$ 
begin
writeln (“рахуємо функції надійності для всіх свідчень:  $Crd(m_i)$ ”);
readln (  $Crd(m_i)$ );
writeln (“рахуємо зважені середні базові маси (коефіцієнти ваги) для всіх гіпотез:  $m_M(A), m_M(B), \dots$ ”);
readln ( $m_M(A), m_M(B), \dots$ );
end.
For  $i:=1$  to  $n-1$  do
begin
writeln (“знаходимо нові базові маси для всіх гіпотез, комбінуючи коефіцієнти ваги за правилом Демпстера:  $m^*(A), m^*(B), \dots$ ”);
readln ( $m^*(A), m^*(B), \dots$ );
end.

```

## 5. Висновки

У даній роботі було розглянуто та проаналізовано декілька методів об'єднання даних, отриманих з різних джерел. Було показано, що правило комбінації Демпстера не можна застосовувати за наявності занадто суперечливих джерел інформації, оскільки воно дає невірні результати. Було описано альтернативний метод усереднення Мерфі, який може оперувати суперечливими даними та давати кращі результати, ніж правило комбінації Демпстера. Але також було зазначено, що і метод Мерфі має певний недолік, оскільки не враховує асоціативний зв'язок між свідченнями.

Тому в даній роботі було запропоновано новий метод комбінування даних, який відображає асоціативний взаємозв'язок між свідченнями, враховує відстань між свідченнями і може ефективно оперувати за наявності занадто суперечливих свідчень [18–20].

У майбутньому запропонований метод об'єднання даних, отриманих із суперечливих джерел, буде використовуватися при розв'язанні задач класифікування лісів, урбанізованих територій, сільськогосподарських земель, при пошуку корисних копалин та розв'язанні різноманітних природоресурсних задач.

## СПИСОК ДЖЕРЕЛ

1. Hegarat-Mascle S. Application of Dempster-Shafer Evidence Theory to Unsupervised Classification in Multisource Remote Sensing / S. Hegarat-Mascle, I. Bloch, D. Vidal-Madjar // IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. – 1997. – Vol. 35, N 4. – P. 1018 – 1031.
2. Shafer G. A Mathematical Theory of Evidence / Shafer G. – Princeton, NJ: Princeton University Press, 1976. – P. 875 – 883.
3. Попов М.О. Сучасні погляди на інтерпретацію даних аерокосмічного дистанційного зондування Землі / М.О. Попов // Космічна наука і технологія. – 2002. – Т. 8, № 2/3. – С. 110 – 115.
4. McCoy R.M. Fields Methods in Remote Sensing / McCoy R.M. – New York: Guilford Press, 2005. – P. 150 – 160.
5. Чандра А.М. Дистанционное зондирование и геоинформационные системы / А.М. Чандра, С.К. Гош. – М.: Техносфера, 2008. – 312 с.
6. Еремеев В. Современные гиперспектральные сенсоры и методы обработки гиперспектральных данных / В. Еремеев, И. Мордвинцев, Н. Платонов // Исследование Земли из космоса. – 2003. – № 6. – С. 80 – 90.
7. Альперт С.І. Новий модифікований метод класифікування гіперспектральних космічних зображень на основі теорії Демпстера-Шейфера / С.І. Альперт // Зб. праць XV-ої Міжнар. молодіжної наук.-практ. конф. «Історія розвитку науки, техніки та освіти» за темою «Гуманістичний зміст мегатехнологічного світу», (Київ, 13 квітня 2017 р.). – Київ, 2017. – С. 97 – 99.
8. Гарбук С. Космические системы дистанционного зондирования Земли / С. Гарбук, В. Гершензон. – М.: Изд-во А и Б, 1997. – 296 с.
9. Попов М. Методы оптимизации числа спектральных каналов в задачах обработки и анализа данных дистанционного зондирования Земли / М. Попов, С. Станкевич // Современные проблемы дистанционного зондирования земли из космоса. – М.: ИКИ РАН, 2006. – Т. 2, № 1. – С. 61 – 63.
10. Аковецкий В.И. Дешифрирование снимков / Аковецкий В.И. – М.: Недра, 1983. – 320 с.
11. Bongasser M. Hyperspectral Remote Sensing: Principles and Applications / M. Borengasser, W.S. Hungate, R. Watkins. – Boca Raton, FL: CRC Press, 2008. – 119 p.
12. Альперт С.І. Новий удосконалений підхід до комбінування даних на основі теорії Демпстера-Шейфера / С.І. Альперт // Зб. матеріалів VII-ої Всеукр. молодіжної наук. конф. «Ідеї та новації в системі наук про Землю», (Київ, 25–27 жовтня 2017 р.). – Київ, 2017. – С. 26 – 27.
13. Chang C.-I. Hyperspectral Data Processing: Algorithm Design and Analysis / Chang C.-I. – Hoboken, NJ: John Wiley & Sons, 2013. – 1164 p.
14. Beynon M.J. The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modeling / M.J. Beynon, B. Curry, P. Morgan // Omega: Int. Journal of Management Science. – 2000. – Vol. 28, N 1. – P. 37 – 50.

15. Taroun A. Dempster-Shafer Theory of Evidence: Potential usage for decision making and risk analysis in construction project management / A. Taroun, J.B. Yang // The Built & Human Environment Review. – 2011. – Vol. 4, Special Issue 1. – P. 155 – 166.
16. Корн Г. Справочник по математике / Г. Корн, Т. Корн. – М.: Наука, 1974. – 831 с.
17. Barnett J.A. Calculating Dempster-Shafer plausibility / J.A. Barnett // IEEE Trans. Pattern Anal. Machine Intell. – 1991. – Vol.13. – P. 599 – 602.
18. Renyi A. Probability theory / Renyi A. – Amsterdam: North-Holland Pub. Co, 1970. – 670 p.
19. Dempster A. Upper and lower probabilities induced by a multivalued mapping / A. Dempster // Annals of Mathematical Statistics. – 1967. – Vol. 38, N 2. – P. 325 – 339.
20. Zhang L. Representation, independence and combination of evidence in the Dempster-Shafer Theory of Evidence / L. Zhang // Advances in the Dempster-Shafer Theory of Evidence. Edited by R.R. Yager, J. Kacprzyk, M. Ftdrizzi. – New York: John Wiley & Sons, 1994. – P. 51 – 69.

*Стаття надійшла до редакції 17.04.2018*