



<https://doi.org/10.36023/ujrs.2021.8.3.201>

UDC 528.8:528.7: 528.71:528.4:528.06

## Data combination method in Remote Sensing tasks in case of conflicting information sources

S. I. Alpert

*Scientific Centre for Aerospace Research of the Earth of the Institute of Geological Science of the National Academy of Sciences of Ukraine, 55-B, O. Gonchar str., Kyiv 01054, Ukraine*

Nowadays technologies of UAV-based Remote Sensing are used in different areas, such as: ecological monitoring, agriculture tasks, exploring for minerals, oil and gas, forest monitoring and warfare. Drones provide information more rapidly than piloted aerial vehicles and give images of a very high resolution, sufficiently low cost and high precision.

Let's note, that processing of conflicting information is the most important task in remote sensing. Dempster's rule of data combination is widely used in solution of different remote sensing tasks, because it can process incomplete and vague information. However, Dempster's rule has some disadvantage, it can not deal with highly conflicted data. This rule of data combination yields wrong results, when bodies of evidence highly conflict with each other. That's why it was proposed a data combination method in UAV-based Remote Sensing. This method has several important advantages: simple calculation and high accuracy.

In this paper data combination method based on application of Jaccard coefficient and Dempster's rule of combination is proposed. The described method can deal with conflicting sources of information. This data combination method based on application of evidence theory and Jaccard coefficient takes into consideration the associative relationship of the evidences and can efficiently handle highly conflicting sources of data (spectral bands).

The frequency approach to determine basic probability assignment and formula to determine Jaccard coefficient are described in this paper too. Jaccard coefficient is defined as the size of the intersection divided by the size of the union of the sample sets. Jaccard coefficient measures similarity between finite sets. Some numerical examples of calculation of Jaccard coefficient and basic probability assignments are considered in this work too.

This data combination method based on application of Jaccard coefficient and Dempster's rule of combination can be applied in exploring for minerals, different agricultural, practical and ecological tasks.

**Keywords:** evidence theory, Jaccard coefficient, basic probability assignment, combination rules, conflicting evidence

© S. I. Alpert. 2021

### 1. Introduction

Nowadays UAV (unmanned aerial vehicles) give new opportunities for conducting environmental monitoring and solution different agricultural and ecological problems. UAV give images with a very high resolution and that's why we can get a detail picture of fields, forests, urban territories (McKnight, 2015; Alpert et al., 2021).

Let's note, that procedure of classification of digital images obtained by unmanned aerial vehicles is one of most important and difficult procedures in remote sensing. The most accurate results are provided by supervised classification methods. In this paper an improved combination method in UAV-based Remote Sensing is considered.

A lot of known classification methods use information from different sources (spectral bands). Information fusion is a one of most important procedures

in land-cover classification. The main aim of aggregation of data is to simplify information from different sources (spectral bands).

Combination rules are the special types of aggregation methods for data obtained from different sources of information (spectral bands) that can provide different assessments for the same hypotheses. Let's note that procedure of processing combining of conflicting information is a very difficult problem in UAV-based Remote Sensing tasks. A lot of known combination methods yield illogical results, when bodies of evidence highly conflict with each other, because these methods of combination can't deal with significant conflict in the data.

**The object of this paper** is description of data combination method, when bodies of evidence highly conflict with each other. This method is based on application of Jaccard coefficient and Dempster's rule of combination. It is noted, that this method can deal with conflicting sources of information (spectral bands). It is also shown, that this combination method based on application of Jaccard coefficient, shows the associative

E-mail: [sonyasonet87@gmail.com](mailto:sonyasonet87@gmail.com)  
<https://orcid.org/0000-0002-7284-6502>

relationship of the evidences and can efficiently handle highly conflicting evidences.

In this work Jaccard coefficient is also considered. It is noted, that this coefficient is defined as the size of the intersection divided by the size of the union of the sample sets. Jaccard coefficient measures similarity between finite sample sets.

The frequency approach to determine basic probability assignment is also analyzed.

Some numerical examples of calculation of Jaccard coefficient and basic probability assignments are described in this paper.

## 2. Jaccard coefficient

Jaccard similarity coefficient was introduced by professor of botany and plant physiology Paul Jaccard in 1901. Jaccard coefficient is a coefficient used for comparing the similarity of sample sets. It is used in various sciences, such as: ecology, informatics, geobotany, biology, remote sensing and others (Kosub, 2019).

Jaccard similarity coefficient between two sample sets  $A_k$  and  $A_l$  is defined as:

$$sim_1(A_k, A_l) = \frac{|A_k \cap A_l|}{|A_k \cup A_l|}, \quad (1)$$

where  $|A_k \cap A_l|$  – cardinality of the intersection of sample sets  $A_k$  and  $A_l$ ;  $|A_k \cup A_l|$  – cardinality of the union of sample sets  $A_k$  and  $A_l$ .

Jaccard similarity coefficient is defined as the cardinality of the intersection of sample sets divided by the cardinality of the union of sample sets.

Suppose, set  $A_k$  corresponds to the vector  $u$  and set  $A_l$  corresponds to the vector  $v$ .

Then Jaccard coefficient can also be introduced as a vector similarity function:

$$sim_1(u, v) = \frac{\sum_{k=1}^t u_k v_k}{\sum_{k=1}^t u_k + \sum_{k=1}^t v_k - \sum_{k=1}^t u_k v_k}, \quad (2)$$

where  $u = (u_1, \dots, u_t)$ ,  $v = (v_1, \dots, v_t)$  are vectors.

### Numerical examples of calculation of Jaccard coefficient

1) Let's consider a first example, where Jaccard coefficient between two sample sets  $A_k$  and  $A_l$  is defined:

$$A_k = \{3, 5, 4, 9, 1, 7, 2\};$$

$$A_l = \{6, 4, 7, 2, 8, 1, 3\}.$$

The intersection between two sets  $A_k$  and  $A_l$  is defined:  $A_k \cap A_l = \{3, 4, 1, 7\}$

The union of two sets  $A_k$  and  $A_l$  is defined as:

$$A_k \cup A_l = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Then we should calculate number of elements of the intersection of two sample sets  $A_k$  and  $A_l$  and number of elements of the union of two sample sets  $A_k$  and  $A_l$

Number of elements of the intersection of two sample sets is “4”, number of elements of the union of two sample sets is “9”.

Jaccard coefficient is defined as the number of elements of the intersection of two sample sets divided by the number of elements of the union of two sample sets :

$$sim_2(A_k, A_l) = \frac{|A_k \cap A_l|}{|A_k \cup A_l|} = \frac{4}{9} = 0.444.$$

2) Let's consider a second example.

We should define Jaccard measure for two vectors  $u$  and  $v$ :

$$u = \{1, 1, 2, 3, 0, 1\};$$

$$v = \{0, 0, 1, 0, 2, 2\}.$$

At the first step, we define the sum of all components of two vectors  $u$  and  $v$ :

$$\sum_{k=1}^t u_k = 1 + 1 + 2 + 3 + 0 + 1 = 8;$$

$$\sum_{k=1}^t v_k = 0 + 0 + 1 + 0 + 2 + 2 = 5.$$

At the second step, we should calculate scalar product of vectors  $u$  and  $v$ :

$$\sum_{k=1}^t u_k v_k = [(1 \cdot 0) + (1 \cdot 0) + (2 \cdot 1) + (3 \cdot 0) + (0 \cdot 2) + (1 \cdot 2)] = 4.$$

Then Jaccard coefficient is defined as:

$$sim_1(u, v) = \frac{\sum_{k=1}^t u_k v_k}{\sum_{k=1}^t u_k + \sum_{k=1}^t v_k - \sum_{k=1}^t u_k v_k} = \frac{4}{8 + 5 - 4} = \frac{4}{9} = 0.444.$$

## 3. Main concepts of Dempster-Shafer evidence theory

In Dempster-Shafer theory, with every hypothesis  $A$  ( $A \subset \Omega$ ) there is associated basic probability assignment  $m(A)$ . The  $m(A)$  value represents the degree of belief allocated to the hypothesis  $A$  (Lein, 2003; Mertikas et al., 2001). Basic probability assignment belongs to the interval  $[0, 1]$  and satisfies the next conditions:

$$\sum_{A \subset \Omega} m(A) = 1; \quad (3)$$

$$m(\emptyset) = 0.$$

Let's consider Dempster rule. Suppose one source appointed to the hypothesis  $A$  a basic probability assignment  $m_1$  and the other source independently appointed to the same hypothesis a basic probability assignment  $m_2$ . The rule of Dempster allows combining these basic probability assignments. Then we get such result:

$$m(A_k) = \frac{\sum_{A_{i1} \cap A_{2j} = A_k} m_1(A_{i1}) m_2(A_{2j})}{1 - C}, \quad (4)$$

where

$$C = \sum_{A_{i1} \cap A_{2j} = \emptyset} m_1(A_{i1}) m_2(A_{2j}). \quad (5)$$

The  $C$  value is called the conflict coefficient and shows the degree of conflict among the sources (spectral bands). Let's note, that  $C \in [0, 1]$ , where “0” indicates the absence of contradictory assessments of the sources of information. The more contradictions we have, the closer is the  $C$  value to “1”.

#### 4. Frequency approach to determine basic probability assignment

It is known a lot of approaches to determine basic probability assignments. Let's consider a frequency approach to determine basic probability assignments.

Suppose, the hyperspectral image consists of a set of spectral images:

$$S_k = \{\pi_n, u_{nk}\}_{n=1}^{N_\pi}, \quad k = 1, \dots, K,$$

where  $N_\pi$  – total number of pixels,  $K$  – number of spectral images; the full signal  $u_n$  of the pixel  $\pi_n$  is considered as a vector with components

$$u_{nk} : u_n = \{u_{nk}\}_{k=1}^K.$$

Let's the pixel  $\pi_n$  displays an object of some class. Each of the axes of this spectral space is divided into intervals. Let's note, that number of intervals is equal to the number of classes. We can represent each class in the spectral space by a set of intervals.

Then consider one interval of the one  $l_1$  class.  $M_j$  – total number of pixels, that belong to this interval.

Let's denote, that  $M_1$  – total number of pixels, that belong to the class with the same name ( $l_1$  class);  $M_2$  – total number of pixels that belong to the other classes.

Then basic probability assignments are defined as (Alpert et. al., 2020):

$$m_1 = \frac{M_1}{M_j}, \quad m_2 = \frac{M_2}{M_j}, \quad (6)$$

where  $m_1$  – basic probability assignment, that classified pixel belongs to the class with the same name ( $l_1$  class);  $m_2$  – basic probability assignment, that classified pixel belongs to the other classes.

We conduct this procedure with other spectral axes and conduct classification, using combination rules.

#### 5. Combination method in case of conflicting information sources

Dempster's rule of combination is the most popular rule of combinations (Popov et al., 2015; Gong, 1996), but this rule has some disadvantages. It can't deal with highly conflicted data sources (spectral bands). And correlation among various evidences isn't taken into account by Dempster's rule. That's why we propose an improved data combination method based on application of Jaccard coefficient and basic concepts of Dempster-Shafer evidence theory. This combination method can efficiently handle highly conflicting evidences and show the associative relationship of the evidences. This method is summarized in next steps:

At first step, we define the Distance measure between pairs of spectral bands (Joussleme et al., 2001). The distance is defined between two bodies of evidence  $m_i$  and  $m_j$  by the following formula:

$$d_{BPA}(m_i, m_j) = \sqrt{\frac{1}{2}(\|m_i\|^2 + \|m_j\|^2 - 2\langle m_i, m_j \rangle)}, \quad d_{BPA} \in [0, 1]. \quad (7)$$

$$\langle m_i, m_j \rangle = \sum_{k=1}^{2^n} \sum_{l=1}^{2^n} m_i(A_k) m_j(A_l) \frac{|A_k \cap A_l|}{|A_k \cup A_l|}, \quad k, l = 1, \dots, 2^n; \quad (8)$$

At the second step, the Similarity measure between two bodies of evidence  $m_i$  and  $m_j$  is defined as:

$$sim(m_i, m_j) = 1 - d(m_i, m_j). \quad (9)$$

At the third step, we should build Measure Matrix. Let's note, that we suppose, that the number of bodies of evidence is  $n$ . Then Similarity Measure Matrix is:

$$SMM = \begin{bmatrix} 1 & s_{12} & \dots & s_{1n} \\ s_{21} & 1 & \dots & s_{2n} \\ \dots & \dots & \dots & \dots \\ s_{n1} & s_{n2} & \dots & 1 \end{bmatrix}. \quad (10)$$

At the fourth step, the Support degree of each piece of evidence  $m_i$  is defined as:

$$Sup(m_i) = \sum_{j=1}^n s_{ij}, \quad i = 1, 2, \dots, n. \quad (11)$$

At the fifth step the Credibility degree of each piece of evidence  $m_i$  is calculated as:

$$Crd(m_i) = \frac{sup(m_i)}{\sum_{i=1}^n sup(m_i)}. \quad (12)$$

At the sixth step the weight average mass  $m_M$  of the evidence for the hypothesis  $a$  is defined as:

$$m_M(a) = \sum_{i=1}^n Crd_i(m_i) \times m_i(a). \quad (13)$$

Then we use Dempster's rule to combine the weighted average of the masses  $n - 1$  times, if we have  $n$  pieces of evidence ( $n$  spectral bands).

#### Pseudocode

$n$ : total number of images  $S_i$   
for  $i$ : 1 to  $n$   
for  $j$ : 1 to  $n$

```

S is a vector of images S
begin
write ('Si = ');
readln (Si);
write ('Sj = ');
readln (Sj);
writeln ("define the Distance measure between two
spectral bands Si and Sj: dBPA (mi, mj)");
readln dBPA (mi, mj);
end.
begin
writeln ("define the Similarity measure between two
spectral bands Si and Sj: sim (mi, mj)");
readln (sim (mi, mj));
end.
Program Matrix ("build the Similarity Measure Matrix")
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 ("Enter", n × n, "elements");
for i:=1 to n do
for j:= 1 to n do
begin
write (' sim [' , j, ', ', 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 ("calculate the sum of the elements of the matrix
row and get a Support degree of each piece of
evidence");
end.
For i:= 1 to n
S is a vector of images S;
Sort s
begin
writeln ("calculate the Credibility degree of each piece
of evidence: Crg (mi)");
readln (n);
writeln ("calculate the weight average mass of the
evidence for each hypothesis: mM (A), mM(B),...
readln (mM (A), mM(B),...);
end.
For i:= n – 1 do
begin
writeln ("combine weight average masses, applying
Dempster combination rule and get new basic
probability assignments");
readln (m* (A), m* (B),...)

```

## Conclusions

UAV give new opportunities for solving different remote sensing tasks, such as: plant health assessment, analysis of natural disasters, surveying mines, mapping deposit sites, solution different ecological problems. Nowadays the classification of digital images obtained by drones is one of most difficult tasks in remote sensing. Let's note, that data fusion is a one of most important procedures in land-cover classification. Data fusion can aggregate information from different sources.

Nowadays processing of conflicting data is a very difficult problem in UAV-based Remote Sensing tasks, because known methods of combination can't deal with significant conflict in the sources of information (spectral bands) and yield wrong solutions of problems, when bodies of evidence highly conflict with each other.

A data combination method based on application of Jaccard coefficient and Dempster's rule of combination was considered in this work. It was shown, that this new method can deal with conflicting sources of information (spectral bands) (Popov et al., 2017; Alpert, 2020).

The frequency approach to determine basic probability assignment was described in this paper. Formulas to determine basic probability assignment were analyzed too.

It also was noted, that Jaccard coefficient measures similarity between finite sample sets.

Several numerical examples of calculation of Jaccard coefficient were analyzed in this work.

It is noted, that this described combination method based on application of Jaccard coefficient and Dempster's rule of combination can be applied in solution of different remote sensing tasks, such as: forestry, plant health assessment, solution of different agricultural problems, exploring for oil, gas and minerals, practical and ecological tasks.

## References

- Alpert, S. (2020). A new approach to applying the discount rule in hyperspectral satellite image classification. *Management of Development of Complex Systems*, 43. 76–82. doi.org/10.32347/2412-9933.2020.43.76–82.
- Alpert, M. I., Alpert, S. I. (2020). New methods to determine basic probability assignment and data fusion in Hyperspectral Image Classification. *Proceedings of the XIX-th International Conference on Geoinformatics. Theoretical and Applied Aspects*. 1–5.
- Alpert, M. I., Alpert, S. I. (2021). A new approach to accuracy assessment of land-cover classification in UAV-based Remote Sensing. *Proceedings of the XXth International Conference on Geoinformatics Theoretical and Applied Aspects*. 1–5.
- Jousselme, Anne-Laure. Grenier, Dominic, Bosse, Eloi. (2001). A new distance between two bodies of evidence. *Information Fusion*, 2. 91–101.
- Gong, P. (1996). *Integrated Analysis of Spatial Data from Multiple Sources: Using Evidential Reasoning and Artificial Neural Network Techniques for Geological Mapping*. *Photogrammetric Engineering and Remote Sensing*, 62 (5). 513–523.
- Kosub S. (2019). A note on the triangle inequality for the Jaccard distance. *Pattern Recognition Letters*. 120. 36–38.
- Lein, J. K. (2003). Applying evidential reasoning methods to agricultural land cover classification. *Int. Journal of Remote Sensing*, 24 (21). 4161–4180.
- McKnight, V. (2015). Drone technology and the Fourth Amendment: aerial surveillance precedent and Kyllo do not account for current technology and privacy concerns. *California Western Law Review*, 51. 263.
- Mertikasm P., Zervakism M. E. (2001). Exemplifying the Theory of Evidence in Remote Sensing Image Classification, *Int. Journal of Remote Sensing*. 22 (6). 1081–1095.
- Popov, M., Alpert, S., Podorvan, V., Topolnytskyi, M., Mieshkov, S. (2015). Method of Hyperspectral Satellite Image Classification under Contaminated Training Samples Based on Dempster-Shafer's Paradigm. *Central European Researchers Journal*. 1 (1). 86–97.
- Popov, M. A., Alpert, S. I., Podorvan, V. N. (2017). Satellite image classification method using the Dempster-Shafer approach. *Izvestiya, atmospheric and oceanic. Physics*, 53 (9). 1112–1122. doi: 10.1134/s0001433817090250/

## МЕТОД КОМБІНУВАННЯ ДАНИХ У ЗАДАЧАХ ДИСТАНЦІЙНОГО ЗОНДУВАННЯ ЗЕМЛІ У ВИПАДКУ СУПЕРЕЧЛИВИХ ДЖЕРЕЛ ІНФОРМАЦІЇ

С. І. Альперт

*ДУ “Науковий центр аерокосмічних досліджень Землі ІГН НАН України”, вул. Олеся Гончара, 55-Б, Київ 01054, Україна*

На сьогодні технології дистанційного зондування на основі БПЛА використовуються у різноманітних сферах, таких як: екологічний моніторинг, сільськогосподарські задачі, пошук корисних копалин, нафти і газу, моніторинг лісів, військова справа. Дрони надають інформацію швидше, ніж пілотовані літальні апарати, зображення дуже високої роздільної здатності, досить низької вартості й високої точності.

Зазначимо, що обробка суперечливої інформації є найбільш важливим завданням у дистанційному зондуванні. Правило комбінування даних Демпстера широко використовується при вирішенні різних завдань дистанційного зондування, оскільки воно може обробляти неповну та неточну інформацію. Але правило Демпстера має певний недолік, воно не може працювати з досить суперечливими даними. Це правило комбінування даних надає невірні результати, якщо частини свідчення є досить суперечливими. Тому запропоновано метод комбінування даних з використанням дистанційного зондування на основі БПЛА. Цей метод має ряд важливих переваг: прості розрахунки та висока точність.

У статті пропонується метод комбінування даних на основі застосування коефіцієнта Жаккара та правила комбінування Демпстера. Описаний метод може працювати із суперечливими джерелами інформації. Метод комбінування даних, заснований на застосуванні теорії свідочств і коефіцієнта Жаккара, враховує асоціативний взаємозв'язок між свідченнями та може ефективно обробляти досить суперечливі джерела даних (спектральні канали).

У статті також описано частотний підхід для визначення базових мас та наведено формулу для визначення коефіцієнта Жаккара. Коефіцієнт Жаккара обчислюється як міра перетину, поділена на міру об'єднання множин. Коефіцієнт Жаккара вимірює подібність між скінченими множинами. У роботі також розглянуті деякі числові приклади розрахунку коефіцієнта Жаккара та базових мас.

Метод комбінування даних, заснований на застосуванні коефіцієнта Жаккара і правилі комбінування Демпстера може застосовуватися при пошуку корисних копалин, в різноманітних сільськогосподарських, практичних та екологічних завданнях.

**Ключові слова:** теорія свідчень, коефіцієнт Жаккара, базові маси, правила комбінування, суперечливі свідчення.

## МЕТОД КОМБИНИРОВАНИЯ ДАННЫХ В ЗАДАЧАХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ В СЛУЧАЕ ПРОТИВОРЕЧИВЫХ ИСТОЧНИКОВ ИНФОРМАЦИИ

С. И. Альперт

*ГУ «Научный центр аэрокосмических исследований Земли ИГН НАН Украины», ул. Олеся Гончара, 55- Б, Киев 01054, Украина*

В настоящее время технологии дистанционного зондирования на основе БПЛА используются в разных сферах, таких как: экологический мониторинг, сельскохозяйственные задачи, поиск полезных ископаемых, нефти и газа, мониторинг лесов и в военном деле. Дроны предоставляют информацию быстрее, чем пилотируемые летательные аппараты и предоставляют изображения с очень высоким разрешением, достаточно низкой стоимостью и высокой точностью.

Отметим, что обработка противоречивой информации является наиболее важной задачей в дистанционном зондировании. Правило комбинирования данных Демпстера широко используется при решении различных задач дистанционного зондирования, так как оно может обрабатывать неполную и неточную информацию. Но правило Демпстера имеет некоторый недостаток, оно не может работать с достаточно противоречивыми данными. Это правило комбинирования данных выдает неверные результаты, когда части свидетельства сильно противоречат друг другу. Вот почему был предложен метод комбинирования данных с использованием дистанционного зондирования на основе БПЛА. У метода несколько важных преимуществ: простые расчеты и высокая точность.

В статье предлагается метод для комбинирования данных на основе применения коэффициента Жаккара и правила комбинирования Демпстера. Описанный метод комбинирования данных может иметь дело с противоречивыми источниками информации. Метод комбинирования, основанный на применении теории свидетельств и коэффициента Жаккара, учитывает ассоциативную взаимосвязь между свидетельствами и может эффективно обрабатывать достаточно противоречивые источники данных (спектральные каналы).

В статье также описан частотный подход для определения базовых масс и приведена формула для определения коэффициента Жаккара. Коэффициент Жаккара вычисляется как мера пересечения, деленная на меру объединения множеств. Коэффициент Жаккара измеряет подобие между конечными множествами. В работе также рассмотрены некоторые числовые примеры расчета коэффициента Жаккара и базовых масс.

Метод комбинирования данных, основанный на применении коэффициента Жаккара и правиле комбинирования Демпстера может применяться при поиске полезных ископаемых, в различных сельскохозяйственных, практических и экологических заданиях.

**Ключевые слова:** теория свидетельств, коэффициент Жаккара, базовые массы, правила комбинирования, противоречивые свидетельства.

*Статья поступила в редакцию 20.07.2021*