

УДК 528.06

Порівняння нового удосконаленого підходу комбінування суперечливих даних з правилом Ягера

С. І. Альперт *

ДУ “Науковий центр аерокосмічних досліджень Землі ІГН НАН України”, Київ, Україна

Теорія свідчень Демпстера-Шейфера є найбільш ефективним підходом для обробки неточної та неповної інформації. Але правило комбінації Демпстера не може працювати з суперечливими даними, що може призвести до невірних результатів. Комбінування суперечливих частин свідчення є однією з найбільш складних задач. Підхід Ягера може обробляти суперечливі дані, але даний підхід не враховує взаємозв'язок між свідченнями. Запропоновано новий метод комбінування, який враховує взаємозв'язок між різними частинами свідчення та дає точні результати.

Ключові слова: правило комбінації Демпстера, теорія свідчень, правило Ягера, суперечливі свідчення

© С. І. Альперт. 2018

Вступ

На даний час дистанційне зондування Землі грає неабияку роль у вивченні стану рослинного покриву, пошуку корисних копалин, оцінці та контролю змін, які виникають у природному середовищі. Класифікування супутникових зображень є одним з надійних методів отримання достовірної інформації про основні характеристики земляного покриву [4–5]. Існує багато різноманітних методів класифікування даних. У статті розглянуто метод, який працює за наявності суперечливих джерел інформації, а саме метод Ягера. Запропоновано новий удосконалений метод комбінування даних за наявності суперечливих джерел інформації, який, на відміну від інших методів, враховує кореляцію між різними джерелами інформації та ефективно вирішує задачу комбінування даних.

Основні положення теорії свідчень Демпстера-Шейфера

Теорія свідчень є певним узагальненням теорії ймовірностей, її математичний апарат дозволяє оперувати не тільки з окремими подіями синглтонами або гіпотезами, але й зі сполученнями подій (гіпотез), причому кожне таке сполучення розглядається за методологією теорії Демпстера-Шейфера (ТДШ) як окрема нерозкладна подія (гіпотеза) A . У ТДШ ключовим поняттям є поняття “маси”, яке є узагальненням класичного поняття ймовірності. На відміну від байєсівської ймовірності, “маса” може ефективно описати незнання та відокремити поняття відсутності довіри від недовіри. “Маса” є мірою довіри до пов'язаної з нею гіпотези [1–4].

Сукупність вихідних (базових) гіпотез відносно стану об'єкта та всі можливі їх сполучення утворюють множину Ω , яка називається у ТДШ “основною аналізу” (*frame of discernment*). Вважається, що всі гіпотези з множини Ω (як базові, так і їхні сполучення) є взаємно незалежними та у сукупності повністю характеризують об'єкт. Якщо число базових гіпотез рівно Q , то загальна кількість підмножин у множині Ω складає величину 2^Q (сюди входять пуста множина \emptyset та сама множина Ω).

Кожному елементу множини Ω надається, за певними правилами, відповідна маса m , причому область значень маси лежить в інтервалі від нуля до одиниці. Формально це записується таким чином:

- маса пустої множини дорівнює нулю: $m(\emptyset) = 0$;
- сума усіх мас для кожної підмножини $A \subseteq \Omega$ дорівнює одиниці: $\sum_{A \subseteq \Omega} m(A) = 1$.

На змістовному рівні масу m можна розглядати як міру довіри до пов'язаної з нею гіпотези; тому $m(A)$ ще називають базовою ймовірністю, а будь-яку підмножину A , для якої $m(A) > 0$, називають фокальним елементом (*focal set*).

Гіпотези, або свідчення можуть одночасно формуватися кількома різними джерелами. У випадку, коли ці джерела взаємно незалежні, в ТДШ розроблено просте правило об'єднання (комбінування) свідчень. Припустимо, ми маємо множину — основу аналізу Ω з числом свідчень (гіпотез) Q , а також маємо P незалежних джерел ID_p ($1 \leq p \leq P$) цих свідчень. При цьому p — те джерело дає свою оцінку ймовірності (можливості) свідчення A_n у вигляді відповідної маси $m_p(A_n)$ або оцінку ймовірності (можливості) сполучення свідчень X у вигляді $m_p(X)$. Тоді об'єднання свідчень, або отримання комбінованого свідчення здійснюється за допомогою правила Демпстера:

* E-mail: sonyasonet87@gmail.com.
Тел.: + 030 95 901 57 37

$$m_D(X) = \frac{1}{1-K} \sum_{\substack{X_1 \cap \dots \cap X_R = X \\ \forall (X \neq \emptyset) \in 2^Q}} \prod_{1 \leq p \leq P} m_p(X_r) \quad (1)$$

де K — коефіцієнт конфліктності,

$$K = \sum_{\substack{X_1 \cap \dots \cap X_R = \emptyset \\ \forall (X \neq \emptyset) \in 2^Q}} \prod_{1 \leq p \leq P} m_p(X_r) \quad (2)$$

Коефіцієнт конфліктності K вказує, наскільки протилежними між собою є джерела (точніше, їх свідчення). Область значень коефіцієнта K знаходиться в інтервалі $[0; 1]$, при цьому нульове значення свідчить про відсутність протиріч в оцінках джерел, а чим сильніше ці протиріччя, тим ближче до одиниці стає величина K .

Ще слід враховувати, що в розрахунках за правилом Демпстера (1) приймається, що $m_p(\emptyset) = 0$.

Метод Демпстера-Шейфера використовується при побудові інтервальних оцінок достовірності гіпотез в умовах відсутності частини вхідних даних, при наявності неповної та неточної інформації, дає правило для комбінування даних від різних джерел, використовується при класифікації гіперспектральних зображень [13–17].

Зазначимо основні недоліки підходу Демпстера-Шейфера:

- 1) Правило об'єднання функцій довіри отримано при припущенні однакової надійності та достовірності різних джерел свідчень. Насправді інформація, що отримується з різних джерел, має різну ступінь достовірності, тобто різну цінність. Це може призводити до помилок. Тому треба визначати найбільш достовірні джерела свідчень.
- 2) Правило комбінації Демпстера може давати не досить точні результати, якщо коефіцієнт конфліктності приймає велике значення, тобто, коли між свідченнями (джерелами інформації) є досить великі протиріччя. Внаслідок цього може виникнути така ситуація, що події, яка в дійсності має найменшу вірогідність, може бути надана найбільша базова маса.

Ось чому дуже важливою є задача модифікації правила Демпстера у випадку суперечливих свідчень [7–12].

Правило Ягера

Розглянемо модифікацію правила комбінації Демпстера — правило Ягера, яке так само як і Демпстер припускає, що $m(\emptyset) = 0$ та визначається наступним чином:

$$m(A) = \sum_{B_1 \cap B_2 \cap \dots \cap B_n = A \neq \emptyset, \Theta} \prod_{1 \leq i \leq n} m_i(B_i), \quad (3)$$

де Θ — основа аналізу (базова множина, фрейм розрізнення), сукупність вихідних гіпотез відносно стану об'єкта та всі можливі їх сполучення.

$$m(\Theta) = \sum_{B_1 \cap B_2 \cap \dots \cap B_n = \Theta} \prod_{1 \leq i \leq n} m_i(B_i) + K, \quad (4)$$

де

$$K = \sum_{B_1 \cap B_2 \cap \dots \cap B_n = \emptyset} \prod_{1 \leq i \leq n} m_i(B_i). \quad (5)$$

Тобто ідея правила Ягера полягає у наданні маси перетинів конфліктних множин (гіпотез), що в перетині дають пусту множину базовій множині. Тобто ненульова маса пустої множини в основному розподіляється серед елементів базової множини (фрейму розрізнення). K означає масу, яка надається базовій множині після процедури комбінування. Правило Ягера використовує інформацію про конфлікт та незнання тільки при обчисленні базових мас множини Θ . Цей факт говорить про “обережність” даного правила. При комбінуванні неконфліктних свідчень, коли $K = 0$ правило Ягера буде співпадати з правилом Демпстера.

Описане правило Ягера може працювати із суперечливими джерелами інформації, але воно не враховує кореляцію між різними джерелами інформації, тобто не враховує асоціативний зв'язок між свідченнями, отриманих з різних джерел.

Тому, у даному випадку пропонується використовувати новий удосконалений метод комбінування даних за наявності великого значення коефіцієнта конфліктності на основі теорії свідчень Демпстера-Шейфера [6–7].

Новий удосконалений метод комбінування даних за наявності великого значення коефіцієнта конфліктності

Представлений метод складається з семи етапів:

- 1) Визначаємо відстань між двома частинами свідчення m_i та m_j :

$$d_{BPA}(m_i, m_j) = \sqrt{\frac{1}{2}(\|m_i\|^2 + \|m_j\|^2 - 2\langle m_i, m_j \rangle)}, \quad d_{BPA} \in [0, 1]. \quad (6)$$

$$\langle m_i, m_j \rangle = \sum_{k=1}^{2^n} \sum_{l=1}^{2^n} m_i(A_k) m_j(A_l) \frac{|A_k \cap A_l|}{|A_k \cup A_l|}, \quad k, l = 1, \dots, 2^n; \quad (7)$$

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 в 1901 році.

У формулі (7) 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 та m_j :

$$\text{sim}(m_i, m_j) = 1 - d(m_i, m_j). \quad (8)$$

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

$$SMM = \begin{bmatrix} 1 & s_{12} & \dots & s_{1n} \\ s_{21} & 1 & \dots & s_{2n} \\ \dots & \dots & \dots & \dots \\ s_{n1} & s_{n2} & \dots & 1 \end{bmatrix}. \quad (9)$$

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

$$\text{Sup}(m_i) = \sum_{j=1}^n s_{ij}, \quad i = 1, 2, \dots, n. \quad (10)$$

5) Здійснюючи об'єднання даних, ми прагнемо дізнатися, яке саме свідчення (спектральний канал), який 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 відрізнитися за ступенем надійності. Ось чому, процедура надання ваги свідченню є досить важливою, 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 наступний: якщо певна частина свідчення знаходить підтримку у інших свідчень, 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, ми опишемо “ступінь надійності” свідчення, яке є вагою та відображає важливість кожної частини свідчення.

6) Надійність кожної частини свідчення визначається, як:

$$\text{Crd}(m_i) = \frac{\text{sup}(m_i)}{\sum_{i=1}^n \text{sup}(m_i)}. \quad (11)$$

Звідси також маємо: $\sum_{i=1}^n \text{Crd}_i = 1$. Слід зазначити,

що, чим коротша відстань від даної частини свідчення до інших частин свідчення, тим більшою є ступінь підтримки (функція підтримки) даної частини свідчення. Якщо частина свідчення отримує значну підтримку у інших частин свідчення, то її ступінь надійності є високою і, вона буде мати більший вплив на кінцевий результат об'єднання даних. Навпроти, якщо певна частина свідчення суперечить іншим свідченням, то її ступінь надій-

ності — низький, отже, ця частина свідчення має значно менший вплив на кінцевий результат комбінації даних отриманих з різних джерел. Таким чином, ступенем надійності є вага, яка є індикатором важливості даного свідчення.

7) Коефіцієнт ваги m_M свідчення для певної гіпотези a :

$$m_M(a) = \sum_{i=1}^n \text{Crd}_i(m_i) \times m_i(a). \quad (12)$$

8) Далі, використовуючи правило Демпстера, комбінуємо коефіцієнти ваги (12) $n-1$ раз за умови, якщо ми маємо n частин свідчення [1–3].

Висновки

Правило комбінації Демпстера не можна застосовувати за наявності занадто суперечливих джерел інформації, оскільки воно дає невірні результати. Альтернативний підхід Ягера полягає у наданні базової маси перетинів конфліктних гіпотез базовій множині. При цьому ненульова маса пустої множини розподіляється серед елементів базової множини (фрейму розрізнення). Коли коефіцієнт конфліктності — нуль, правило Ягера співпадає з правилом Демпстера.

Правило Ягера може працювати із суперечливими джерелами інформації, але воно не враховує кореляцію між різними джерелами інформації, тобто не враховує асоціативний зв'язок між свідченнями, отриманих з різних джерел.

Запропоновано новий метод комбінування даних, який відображає асоціативний взаємозв'язок між свідченнями, враховує відстань між свідченнями та може ефективно оперувати за наявності занадто суперечливих свідчень.

Метод об'єднання даних, отриманих з суперечливих джерел може використовуватися при розв'язанні різноманітних природноресурсних задач таких як: класифікування лісів, урбанізованих територій, сільськогосподарських земель, при пошуку корисних копалин, родовищ нафти та газу [6–8].

Література

1. Аковецкий В. И. Дешифрирование снимков / В. И. Аковецкий. — М.: Недра, 1983. — 320 с.
2. Альперт С. І. Оцінка точності класифікації космічних зображень на основі теорії Демпстера-Шафера / С. І. Альперт // Збірник праць XI-ої Міжнародної молодіжної науково-практичної конференції “Історія розвитку науки, техніки та освіти” за темою “Розбудова дослідницького університету”, (Київ, 25 квітня 2013 р.). — Київ, 2013. — С. 242–245.
3. Альперт С. І. Новий удосконалений підхід до комбінування даних на основі теорії Демпстера-Шейфера / С. І. Альперт // Збірник матеріалів VII-ої Всеукраїнської молодіжної наукової конференції “Ідеї та новації в

- системі наук про Землю”, (Київ, 25–27 жовтня 2017). — К., 2017. — С. 26–27.
4. Кочуб Е. В. Анализ методов обработки материалов дистанционного зондирования Земли / Е. В. Кочуб, А. А.Топаз // Вестник Полоцкого государственного университета, 2012. — 132 с.
 5. Кузнецов А. В. Сравнение алгоритмов управляемой элементной классификации гиперспектральных изображений / А. В. Кузнецов, В. В. Мясников // Компьютерная оптика. — 2014. — Т. 38. — №3. — С. 494 – 502.
 6. Лурье И. К. Теория и практика цифровой обработки изображений / И. К. Лурье, А. Г. Косиков. — М.: Научный мир, 2003. — 356 с.
 7. Лялько В. И. Использование данных спутниковой съемки для изучения природоресурсных проблем / В. И. Лялько, А. Д. Федоровский, М. А. Попов [и др.] / Космічні дослідження в Україні 2002–2004. — К.: 2004. — С. 2–14.
 8. Марченко Н. А. Методика классификации объектов по данным дистанционного зондирования Земли / Н. А. Марченко, А. К. Авраменко // Системный анализ, управление и информационные технологии. Научное периодическое издание. — Харьков: Харьковский политехнический институт. — 2008. — Вып. №26. — С. 145–149.
 9. Попов М. Підхід до класифікації стану лісових ресурсів по багатоспектральних космічних зображеннях на основі принципу злиття даних та використання елементів ГС-технологій / М. Попов, О. Рябоконтенко, О. Петроченко // Ученые записки Таврического национального университета им. В. И. Вернадского, сер. География. — 2003. — Т. 16 (55). — №2. — С. 142–150.
 10. 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. — 2006. — Т. 2. — №1. — С. 61–63.
 11. Попов М. Шляхи отримання космічної інформації в інтересах національної безпеки та оборони / М. Попов // Наука і оборона. — 2002. — №2. — С. 38–50.
 12. Chang C. I. Hyperspectral Data Processing: Algorithm Design and Analysis / C. I. Chang // Hoboken, NJ: John Wiley & Sons, 2013. — 1164 p.
 13. Gong P. Integrated Analysis of Spatial Data from Multiple Sources: Using Evidential Reasoning and Artificial Neural Network Techniques for Geological Mapping / P. Gong // Photogrammetric Engineering and Remote Sensing. — 1996. — Vol. 62. — № 5. — P. 513–523.
 14. Lein J. K. Applying evidential reasoning methods to agricultural land cover classification / J. K. Lein // Int. Journal of Remote Sensing. — 2003. — Vol. 24. — № 21. — P. 4161–4180.
 15. Mertikas P. Exemplifying the Theory of Evidence in Remote Sensing Image Classification / P. Mertikas, M. E. Zervakis // Int. Journal of Remote Sensing. — 2001. — Vol. 22. — № 6. — P. 108–1095.
 16. Popov M. A. A Dempster-Shafer evidence theory-based approach to object classification on multispectral / hyperspectral images // M. A. Popov, M. V. Topolnitskiy // Proceedings of the 10th International Conference IEEE on Digital Technologies (DT'2014). — Zilina, Slovakia, 9–11 July 2014. — P. 296–300.
 17. Smets P. Constructing the pignistic probability function in a context of uncertainty / P. Smets, M. Henrion, R. D. Shachter, L. N. Kanal, J. F. Lemmer // Uncertainty in Artificial Intelligence. — North Holland, Amsterdam, 1990. — Vol. 5. — P. 29–40.

СРАВНЕНИЕ НОВОГО УСОВЕРШЕНСТВОВААННОГО ПОДХОДА КОМБИНИРОВАНИЯ ПРОТИВОРЕЧИВЫХ ДАННЫХ С ПРАВИЛОМ ЯГЕРА

С. И. Альтерт

Теория свидетельств Демпстера-Шейфера является наиболее эффективным подходом для обработки неточной и неполной информации. Но правило комбинации Демпстера не может работать с противоречивыми данными, что может привести к неверным результатам. Вот почему комбинирование противоречивых частей свидетельства является одной из наиболее сложных задач. Подход Ягера может обрабатывать противоречивые данные, но данный подход не учитывает взаимосвязь между свидетельствами. Предложен новый метод комбинирования, который учитывает взаимосвязь между разными частями свидетельства и дает точные результаты.

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

COMPARISON OF THE NEW IMPROVED COMBINATION APPROACH THAT DEAL WITH CONFLICT DATA WITH YAGER'S SOLUTION APPROACH

S. I. Alpert

Dempster-Shafer evidence theory is the most effective approach to process imprecise and incomplete information. But Dempster's combination rule can not deal with conflicting data and can lead to illogical results. That's why the combination of conflicting bodies of evidence is one of the most difficult problems. Yager's solution approach can process conflicting data, but correlation among evidences is not taken into account. It was proposed a new combination method, that uses correlation among different bodies of evidence and give accurate results.

Keywords: Dempster's combination rule, evidence theory, Yager's solution approach, conflicting bodies of evidence.