

УДК 528.854:519.722:004.93:004.62

М.О. ПОПОВ*, Л.М. АРТЮШИН**, М.В. ТОПОЛЬНИЦЬКИЙ***, О.В. ТИТАРЕНКО*,
Ю.В. ШМОРГУН****

ІНТЕГРУВАННЯ ГЕТЕРОГЕННИХ ГЕОПРОСТОРОВИХ ДАНИХ НА ОСНОВІ ТЕОРІЇ СВІДЧЕНЬ ДЕМПСТЕРА-ШЕЙФЕРА

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

**Державний науково-дослідний інститут авіації, м. Київ, Україна

***Воєнно-дипломатична академія імені Євгенія Березняка, м. Київ, Україна

****Київський національний університет імені Тараса Шевченка, м. Київ, Україна

Анотація. У статті розглядається підхід до інтегрування гетерогенних геопросторових даних на базі єдиної науково-методичної платформи – теорії свідчень Демпстера-Шейфера. Наведено структурно-логічну схему інтегрування гетерогенних геопросторових даних, суть якої зводиться до послідовного інтегрування таких даних на трьох базових рівнях: фізичному, логіко-семантичному та прийняття рішень. Проаналізовано особливості інтегрування гетерогенних геопросторових даних на фізичному, логіко-семантичному рівнях та специфіку їх інтегрування на рівні гіпотез (рішень) у рамках теорії свідчень Демпстера-Шейфера, яка полягає у необхідності релевантного опису конфлікту, що потенційно може виникнути при наявності суперечливих свідчень (гіпотез) та отриманні інтервальної оцінки ймовірності настання події, що у свою чергу передбачає розробки або обрання певних правил, які дозволяють її точно розрахувати у визначеному інтервалі. Крім того, приведено загальні обмеження інтегрування та сформульовано нові підходи до оцінки ефективності виконання тематичних завдань, які зводяться до обрахунку рівня невизначеності на трьох різних рівнях: інформаційних джерел, скомбінованих (інтегрованих) розподілів гіпотез та прийняття рішень. Запропоновано структурно-логічну схему нового методу, який дає змогу враховувати корельованість і узгодженість гіпотез, конфліктність джерел геопросторових даних, та представлено приклади інтегрування геопросторових даних класичним (комбінаційне правило Демпстера) і авторським методами при вирішенні задач пошуку родовищ вуглеводнів та ідентифікації компактних штучних об'єктів. Підтверджено ефективність запропонованого методу, визначено умови і обмеження його застосування.

Ключові слова: інтегрування, геопросторові дані, інтероперабельність, невизначеність, узгодженість гіпотез, конфліктність джерел, теорія свідчень Демпстера-Шейфера.

Аннотация. В статье рассматривается подход к интегрированию гетерогенных геопространственных данных на базе единой научно-методической платформы – теории свидетельств Демпстера-Шейфера. Приведена структурно-логическая схема интегрирования гетерогенных геопространственных данных, суть которой сводится к последовательному интегрированию таких данных на трех базовых уровнях: физическом, логико-семантическом и принятия решений. Проанализированы особенности интегрирования гетерогенных геопространственных данных на физическом, логико-семантическом уровнях, а также специфика их интегрирования на уровне гипотез (решений) в рамках теории свидетельств Демпстера-Шейфера, которая заключается в необходимости релевантного описания конфликта, который потенциально может возникнуть при наличии противоречивых свидетельств (гипотез) и получении интервальной оценки вероятности наступления события, что в свою очередь предусматривает разработки или выбор определенных правил, которые позволяют ее точно рассчитать в определенном интервале. Кроме того, приведены общие ограничения интегрирования, сформулированы новые подходы к оценке эффективности выполнения тематических заданий, которые сводятся к определению уровня неопределенности на трех разных уровнях: информационных источников, скомбинированных (интегрированных) распределений гипотез и принятия решений. Предложена структурно-логическая схема нового метода, который дает возможность учитывать коррелированность и согласованность гипотез, конфликтность источников геопространственных данных, а также представлены примеры интегрирования геопространственных данных классическим (комбинационное правило Демпстера) и

авторским методами при решении задач поиска месторождений углеводородов, идентификации компактных искусственных объектов. Подтверждена эффективность предложенного метода, определены условия и ограничения его применения.

Ключевые слова: интеграция, геопространственные данные, интероперабельность, неопределенность, согласованность гипотез, конфликтность источников, теория свидетельств Демпстера-Шейфера.

Abstract. The article deals with an approach towards integration of geospatial data based on the unified research and methodical platform – the Dempster-Shafer theory of evidence. The article presents a structural chart of heterogeneous geospatial data integration which enables a consecutive integration of such data at three basic levels: physical, logical-semantic and decision-making. It analyzes peculiarities of heterogeneous geospatial data integration at physical and logical-semantic levels as well as particulars of their integration at the level of hypotheses (decisions) within the Dempster-Shafer theory of evidence which envisages a relevant conflict description that can potentially break out in case of controversial evidences (hypotheses) as well as obtaining interval evaluation of the event likelihood of occurrence, which, in its turn, requires development and selection of certain rules allowing its precise calculation in the determined interval. Besides, the article sets forth general limitations of integration and new approaches to efficiency of topical tasks evaluation that basically include calculation of the uncertainty levels at three different levels: information sources, combined (integrated) distributions of hypotheses and decision-making. The article presents a structural-logical chart of a new approach enabling to consider hypotheses correlation and coordination, conflicts of geospatial data sources. It also provides examples of geospatial data integration to classical (the Dempster's combination rule) and author's methods during search for hydrocarbon deposits and identification of compact artificial objects. The article proves the efficiency of the proposed method. It defines conditions and restrictions for its use.

Keywords: integration, geospatial data, interoperability, uncertainty, hypotheses coordination, conflicts of sources, Dempster-Shafer theory of evidence.

DOI: 10.34121/1028-9763-2019-3-35-46

1. Вступ

Як свідчить досвід вирішення природоресурсних, сільськогосподарських, наукових, оборонних та інших складних задач, виникає потреба у залученні різноманітної інформації щодо об'єктів інтересу. Це, в залежності від потреб, може бути геологічна, геофізична, оптична, радіолокаційна та інша інформація. Для її отримання використовуються відповідні датчики, сенсори, бази даних. Особливо важливу роль відіграють геопросторові дані (ГПД), які формуються на основі матеріалів аерокосмічної та наземної зйомки, різноманітних карт, GPS-приладів тощо.

У статті представлено системний підхід до інтегрування ГПД, розглянуто структурно-логічну схему інтегрування, особливості проведення інтегрування на різних рівнях, обмеження інтегрування, підходи до оцінки ефективності виконання тематичних завдань та приклади їх вирішення на основі теорії свідчень Демпстера-Шейфера (ТСДШ).

2. Постановка задачі

Вивчення об'єкта інтересу та прийняття рішень щодо його стану або подальшої поведінки потребує наявності єдиної, узгодженої інформаційної моделі. Створюється така модель у результаті аналізу інформації, що отримується шляхом комплексування (інтегрування) гетерогенних даних від різних джерел. Але при цьому існують дві проблеми.

Перша проблема полягає у тому, що датчики геопросторових та інших необхідних даних мають різну фізичну природу, внаслідок чого дані можуть бути кількісними, описовими (номінальними), аналоговими, цифровими тощо. Вони можуть мати різну достовірність і точність, відрізнятись масштабом, вимірністю (2D, 3D), форматом представлення. Тобто, при вирішенні зазначених вище задач неминуче мають справу з гетерогенними даними.

Друга проблема полягає у тому, що трансформування даних в інформацію та її ком-плексування для послідуочого прийняття рішень – багатоетапний логіко-обчислювальний процес, можливості автоматизації якого внаслідок його суттєвої складності досить обмежені. Тому на сучасному етапі розвитку інформаційних технологій цей процес може бути реалізований лише за участю людини.

Інтегрування гетерогенних ГПД в таких умовах неможливо без відповідної науково-методичної платформи.

Метою статті є обґрунтований вибір такої платформи та розроблення на її базі нового підходу до інтегрування гетерогенних ГПД.

3. Аналіз можливих шляхів вирішення задачі

Моделі і методи, які розробляються для вирішення складних науково-прикладних задач, звичайно спираються на математичний інструментарій або теорії ймовірностей і статистичних рішень, або теорії нечітких множин. Кожен із цих інструментаріїв має свої особливості і нішу застосування.

Застосування ймовірнісно-статистичних методів можливе лише за умови наявності достатньо великих вибірок даних, що описують властивості об'єкта дослідження, причому області значень таких даних підпорядковуються певним ймовірнісним розподілам. Ще одним недоліком ймовірнісно-статистичних методів є їх неспроможність обробляти суперечливі дані та дані, відносно яких відсутня будь-яка інформація, тобто фактично існує «незнання» щодо об'єкта дослідження [1].

Основна перевага використання теорії нечітких множин полягає у наявності великої кількості комбінаційних операторів, які дозволяють користувачу адаптувати схему інтеграції із специфікою даних [2]. Проте, вибір таких операторів потребує експертних оцінок. Крім того, ускладнена валідація результатів, які одержуються методами теорії нечітких множин.

Враховуючи недоліки вказаних методів, перспективним є вивчення можливостей інтегрувати гетерогенні ГПД за допомогою інструментарію ТСДШ. Теорія свідчень може бути інтерпретована як узагальнення теорії ймовірності. У теорії ймовірності свідчення пов'язані тільки з однією із можливих подій. У ТСДШ свідчення може бути пов'язано одночасно з кількома можливими подіями – множиною подій. У результаті в ТСДШ вони можуть мати сенс на більш високому рівні абстракції без необхідності вдаватися до припущень щодо свідчення в межах множини. Там, де свідчення достатньо обґрунтовані та дозволяють однозначно визначити їх ймовірність, ТСДШ використовує модель традиційної ймовірності. Однією з її найбільш важливих особливостей є те, що вона дозволяє усунути неоднозначність визначення різного рівня точності у відношенні інформації та виключає необхідність у додаткових припущеннях щодо її подання. Крім того, вона дозволяє визначити пряму невизначеність відгуку системи у випадку, коли неточності на вході та на виході характеризуються множиною або інтервалом [3].

Тому саме ТСДШ пропонується нижче як науково-методична платформа для інтегрування гетерогенних ГПД.

4. Структурно-логічна схема інтегрування гетерогенних геопросторових даних

За схемою, наведеною на рис. 1, інтегрування ГПД має здійснюватися на трьох рівнях: фізичному, логіко-семантичному та гіпотез (рішень). Інтегрування гетерогенних даних на фізичному рівні обумовлено їх різноманіттям і відмінностями та зводиться до конверсії даних в єдиний формат фізичного представлення. Ця задача пов'язана з проблемою інтероперабельності [3]. Вона вирішується шляхом взаємного узгодження багатьох факторів (рис. 2), якими описуються дані. Таке узгодження означає трансформування кожного фак-

тора до деякого уніфікованого вигляду, що дозволяє далі порівнювати їх в єдиній шкалі та інтерпретувати в одній логіко-семантичній системі оцінок.

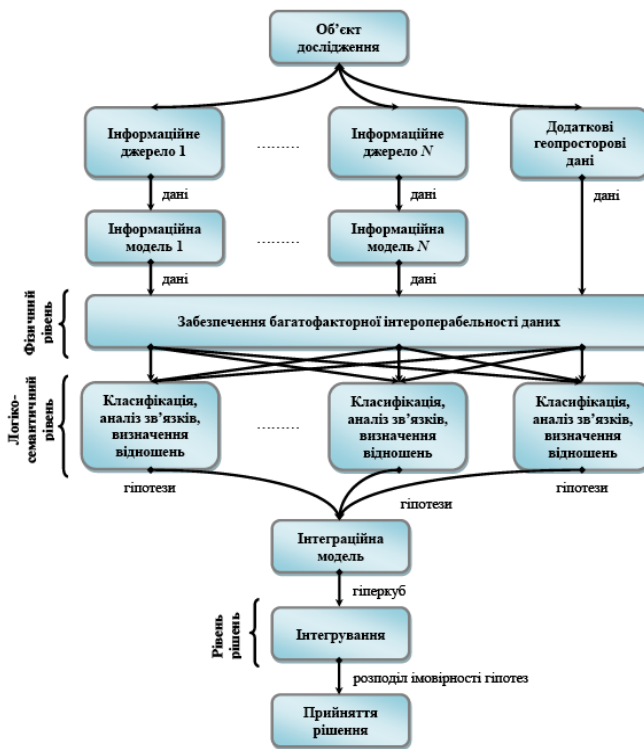


Рисунок 1 – Структурно-логічна схема інтегрування гетерогенних даних



Рисунок 2 – Фактори забезпечення інтероперабельності даних



Рисунок 3 – Послідовність проведення інтегрування ГПД на логіко-семантичному рівні

Інтегрування даних на логіко-семантичному рівні включає селекцію найбільш цінних і характерних властивостей об'єкта досліджень по наборах даних, формування гіпотез щодо його категорії (класу), встановлення зв'язків та відносин у контексті предметної області. Основними етапами проведення інтегрування на цьому рівні є процедури генерування множин гіпотез, їх аналіз і оцінювання, вибір найбільш імовірних гіпотез, а також обрання відповідних методів класифікування, аналізу зв'язків та визначення відношень (рис. 3).

Як вже відзначалось, генерування, аналіз та оцінювання гіпотез – складна інтелектуальна задача, особливо, якщо кожна ситуація має свої особливості і, як правило, не повторюється, а завдання виконуються звичайно в умовах невизначеності, обумовленої чинником випадковості і браком знань. Тому зазвичай генерування множини гіпотез здійснюється безпосередньо людиною. Серед гіпотез, які генеруються, можуть бути як позитивні, так і негативні, тобто такі, які відповідно підтверджують наявність певної властивості у об'єкта інтересу або ні. При

оцінюванні правдоподібності гіпотез застосовують методи дедуктивного [4] і абдуктивного [5].

З точки зору автоматизації когнітивно-аналітичних процесів найбільш перспективним є кількісний підхід до оцінювання ймовірностей гіпотез, однак психологічні дослідження і практика показують, що людина звичайно намагається уникнути жорстких однозначних оцінок (hard estimate), і значно більш схильна до так званих м'яких оцінок (soft estimates) [6].

«Жорстка» гіпотеза характеризується тільки одним твердженням, наприклад, досліджуваний об'єкт – це «штучний об'єкт». «М'яку» гіпотезу можна представити сукупністю кількох тверджень щодо досліджуваного об'єкта (явища), які об'єднані математичним оператором «АБО». Наприклад, досліджуваний об'єкт – це «штучний об'єкт» або «природний об'єкт». Можливість формування таких гіпотез є об'єктивною реальністю під час вивчення будь-яких складних або простих об'єктів (явищ), оскільки на практиці вкрай рідко зустрічають дані, аналіз яких дозволяє з високою достовірністю висувати «жорсткі» гіпотези. Така особливість робить сам процес дослідження об'єкта (явища) більш гнучким і дозволяє уникнути значних помилок під час його аналізу та інтерпретації.

Результатом оцінювання гіпотез є визначення значення їх імовірності. Тому основною задачею цього етапу є розрахунок зазначених величин для всіх альтернативних гіпотез. Фінальним кроком процедури формування та дослідження гіпотез є вибір множини, яка найбільш адекватно описує (інтерпретує) вхідні дані. Зазначений етап передбачає порівняння різних інтерпретацій вхідних даних і є, в певному сенсі, вирішенням оптимізаційної задачі. Для вирішення цього питання можуть застосовувати методи математичного програмування, ймовірнісні або інші методи. Проведення процедур класифікування, аналізу зв'язків та визначення відношень передбачає обрання відповідних методів та підходів з урахуванням специфіки гетерогенних даних.

Інтегрування на рівні гіпотез (рішень) передбачає об'єднання множин гіпотез (класифікаційних рішень) за певними правилами в єдиний інформаційний універсум. Для вирішення зазначеного питання пропонується застосовувати ТСДШ [7–9].

Головною особливістю під час інтегрування на цьому рівні є питання обґрунтованості конфлікту – його релевантність [3, с. 102]. У загальному випадку вона може бути описана за рахунок урахування особливостей джерел надходження даних, свідчень, які вони формують, та оператора комбінування, що використовується для їх інтегрування (рис. 4).



Рисунок 4 – Особливості інтегрування гетерогенних ГПД на рівні гіпотез (рішень)

Саме інтегрування на рівні гіпотез (рішень) є найбільш відповідальним етапом. На основі отриманих даних на цьому рівні формуються висновки, складається прогноз та приймається рішення щодо досліджуваного об'єкта (явища).

5. Особливості інтегрування гетерогенних геопросторових даних на рівні гіпотез (рішень)

Відповідно до структурно-логічної схеми (див. рис. 1), перед проведенням інтегрування гетерогенних ГПД на рівні гіпотез (рішень) формується інтеграційна модель. Для її побудови можуть бути застосовані методи мультимодельовання, зокрема [10]: стикувальний (docking), груповий (collaboration), взаємодії (interoperation), інтеграційний (integration) та зв'язності (coupling). Після цього відбувається комбінування (інтегрування) сформованих гіпотез (класифікаційних рішень) на основі ТСДШ.

У рамках ТСДШ є три важливі функції: функція базового ймовірнісного розподілу (ВРА) або функція маси (m), функція довіри (Belief – Bel) та функція правдоподібності (Plausibility – Pl). Функція базового ймовірнісного розподілу (функція маси) може бути інтерпретована як класична ймовірність [11]. Така концепція була підтримана в [12]. Це дуже важлива і корисна інтерпретація ТСДШ, проте необхідно зауважити, що вона не завжди може продемонструвати повний спектр репрезентативних можливостей ВРА [13]. Величина ймовірності події (у класичному розумінні) перебуває в межах нижньої і верхньої границь – Belief і Plausibility відповідно:

$$Bel(A) \leq P(A) \leq Pl(A). \quad (1)$$

Вона визначається однозначно, якщо $Bel(A) = P(A)$. У даному випадку це відповідає класичній теорії ймовірності, тобто всі ймовірності $P(X)$ однозначно визначаються для всіх підмножин A універсальної множини Ω [14, с. 97]. В іншому випадку функції $Bel(A)$ і $Pl(A)$ можна розглядати як нижню і верхню границі ймовірності, відповідно, де фактичне значення ймовірності $P(A)$ знаходиться в інтервалі, що описується. Верхня і нижня границі ймовірності, які отримані іншими підходами до загальної теорії інформації, не можуть бути безпосередньо інтерпретовані як функції $Bel(A)$ та $Pl(A)$ [15, с. 216].

Тому для визначення величини ймовірності настання події, яка у подальшому застосовується для прийняття інформаційних рішень, необхідно розробити або обрати відповідні підходи, що дозволяють її точно розрахувати у визначеному інтервалі.

6. Обмеження інтегрування гетерогенних геопросторових даних

Застосування гетерогенних ГПД при вирішенні різноманітних тематичних завдань накладає ряд обмежень на можливість їх сумісного оброблення та аналізу. В роботі [16] наведено ряд таких обмежень. Серед них необхідно виділити такі:

1. Інтегрування будь-якої кількості ГПД не може замінити джерело (сенсор), яке найбільш адаптовано під середовище або об'єкт, що досліджується. Будь-яка комбінація сенсорів може забезпечити надійну роботу системи, проте, окремо зазначені сенсори не дозволяють визначити характеристики об'єкта або явища, що становить інтерес. Д. Холомом (D. Hall) наведений приклад механічної системи, в якій окремі сенсори вимірюють швидкість її роботи, рівень вібрації та тиску, що становить цінну інформацію про експлуатаційні характеристики зазначеної системи, проте, навіть у сукупності, вони не дозволяють виміряти значення її поточної температури.

2. Наступний рівень (етап) оброблення вхідних ГПД не дозволяє компенсувати або виправити помилки попереднього рівня (етапу) оброблення. Інтегрування різнорідних ГПД не дозволяє виправляти помилки попереднього етапу (рівня) оброблення або відсут-

ності попереднього оброблення вхідних даних. Наприклад, нездатність ідентифікувати або отримувати необхідні параметри об'єкта дослідження або явища не можуть бути виправлені за допомогою складних алгоритмів розпізнавання образів.

3. Результат інтегрування гетерогенних ГПД може бути гіршим, ніж надана інформація окремим джерелом (сенсором), якщо використовується некоректна математична модель, яка описує принцип дії сенсора. Застосування некоректних математичних моделей, що описують принцип дії сенсора, призведе до викривлення результату інтегрування даних, оскільки їх вклад або ефект буде невірним.

4. Під час інтегрування гетерогенних ГПД не існує універсальних алгоритмів їх комбінування. Не існує оптимальних алгоритмів інтегрування даних. Зазвичай реальні прикладні програмні продукти не задовольняють обмеженням алгоритмів, які використовуються для інтегрування різнорідних ГПД (наприклад, наявність апріорних даних або статистично незалежних джерел надходження інформації).

5. Під час інтегрування гетерогенних ГПД завжди не вистачає навчальних вибірок. У загальному випадку завжди існує проблема з навчальними вибірками, які застосовуються під час, наприклад, розпізнавання образів. Тому необхідно застосовувати гібридні методи, до яких можна віднести методи моделювання, синтаксичного представлення або їх комбінацію.

6. Важко оцінити результат, який отримується шляхом інтегрування гетерогенних ГПД. Однією із головних задач під час розробки системи інтегрування різнорідних ГПД є оцінка ефективності її застосування в реальних умовах. Навіть при наявності показників ефективності окремих джерел (сенсорів) щодо виконання поставлених завдань або алгоритмів обробки визначити ефективність системи зазвичай складно, оскільки не завжди зрозуміло, яким чином той або інший підхід до інтегрування даних дозволить якісно функціонувати системі в операційному середовищі.

7. Інтегрування гетерогенних ГПД є нестатичним процесом. Процес інтегрування різнорідних ГПД є ітераційним, динамічним процесом, який постійно спрямований на уточнення оцінки об'єкта дослідження.

Безумовно, наведені вище обмеження мають бути враховані під час розроблення відповідних методів або способів інтегрування різнорідних ГПД.

7. Невизначеність як характеристика контенту для прийняття інформаційних рішень

Як було відзначено, одним з обмежень процесу інтегрування різнорідних ГПД є складність оцінки його результату. Це пов'язано, в першу чергу, з тим, що наявні (класичні) показники ефективності, наприклад, достовірність правильного дешифрування (розпізнавання), не можуть бути напряму використані. Таку ситуацію можна пояснити тим, що в результаті інтегрування даних є ситуація, коли існує багатоальтернативність рішення, що приймається. Причому жодна з альтернатив не виключається і має власне значення ваги (достовірності настання події, яку вона характеризує). Крім того, за результатами інтегрування максимальну вагу може мати альтернатива, яка мала мінімальне вхідне значення або може відбутись незначне зменшення ваги гіпотези, яка у первинному розподілі мала максимальне значення і за результатами інтегрування його зберегла. Тому традиційний підхід до оцінки інформаційного контенту, який полягає у застосуванні ймовірно-статистичних методів як єдиний і достатній, не може бути застосований.

З огляду на це, для характеристики інформаційного контенту, на основі якого приймаються інформаційні рішення, авторами пропонується оцінювати рівень невизначеності (T). Такий підхід обумовлений тим, що процес аналізу та інтерпретації інтегрованих даних розглядається як інформаційний процес, у результаті якого знімається деяка невизначеність щодо досліджуваного об'єкта (явища) [17, с 12]. У свою чергу, збільшення кількості інформації або її приріст безпосередньо впливають на достовірність його аналізу та

інтерпретації. При наявності приросту інформації щодо досліджуваного об'єкта потенційно має відбуватись підвищення достовірності його аналізу та інтерпретації [18, 19, с. 43].

Для оцінки ефективності виконання тематичного завдання при інтегруванні гетерогенних ГПД, крім орієнтування на відповідну вагу (достовірність) кожної альтернативи, пропонується оцінювати приріст інформації ΔI m щодо досліджуваного об'єкта (явища), який можна обрахувати за таким виразом:

$$\Delta I(m) = T^{pr}(m) - T^{ps}(m), \quad (2)$$

де $T^{pr}(m)$, $T^{ps}(m)$ – первинний та апостеріорний рівні невизначеності щодо об'єкта (явища) відповідно.

У рамках ТСДШ можна виділити три рівні обрахунку значення невизначеності: рівень інформаційних джерел, скомбінованих (інтегрованих) розподілів гіпотез та прийняття рішень. Для обрахунку значення невизначеності на рівні інформаційних джерел та скомбінованих (інтегрованих) розподілів гіпотез можуть бути застосовані вирази, запропоновані Дж. Кліром (G. Klir) [13, 20]:

– для рівня інформаційних джерел:

$$T_{source}(m) = \sum_{A \in \Omega} m(A) \log_2 \sum_{B \in \Omega} m(B) \frac{|A|^2}{|A \cap B|}, \quad (3)$$

де $m(A)$, $m(B)$ – функція маси для підмножин A і B відповідно;

$|A|$ – кількість елементів (кардинальність) підмножини A ;

$|A \cap B|$ – кількість елементів перетину підмножин A і B .

Такий підхід до обрахунку невизначеності може бути використаний для проведення процедур оптимізації обрання інформаційних джерел ГПД при вирішенні тих чи інших тематичних задач;

– для рівня скомбінованих (інтегрованих) розподілів гіпотез:

$$T_{hypotheses}(m) = - \left(\sum_{A \in \Omega} m(A) \log_2 |A| + \frac{1}{c} \sum_{A \in \Omega} (Bel(A) \log_2 Bel(A) + Pl(A) \log_2 Pl(A)) \right), \quad (4)$$

де c визначається як

$$c = \sum_{A \in \Omega} (Bel(A) + Pl(A)). \quad (5)$$

Для обрахунку невизначеності на рівні прийняття рішень може бути застосовано значення ймовірнісної інформаційної ємності (Probabilistic Information Content (PIC) [21]:

$$PIC = 1 + \frac{1}{\log_2 N} \cdot \sum_{n=1}^N P(A_n) \log_2 (P(A_n)), \quad (6)$$

де N – загальна кількість гіпотез.

Підходи, які представлені виразами (4)–(6), є квінтесенцією вирішення конкретної тематичної задачі при проведенні інтегрування різнорідних ГПД та поряд із використанням класичних імовірнісно-статистичних методів дозволяють додатково оцінити отриманий результат.

Таким чином, можна констатувати, що особливістю при інтегруванні різнорідних ГПД є процедура обрахунку рівня невизначеності щодо досліджуваного об'єкта (явища). Невизначеність при інтегруванні гетерогенних ГПД – це нове поняття для характеристики контенту, яке призначене для осіб, що приймають інформаційні рішення.

8. Приклади вирішення тематичних задач на основі запропонованого підходу

Було застосовано комбінаційне правило Демпстера для інтегрування геолого-геофізичних даних, супутникових даних і окремих шарів цифрової моделі місцевості при вирішенні двох задач: пошуку покладів нафти і газу та ідентифікації компактних штучних об'єктів. Результати інтегрування наведено на рис. 5 та рис. 6.

Подібний підхід дозволив:

1. Спрогнозувати нафтогазоносність для території Липоводолинського родовища. Проте, для більшої її частини такий прогноз став неможливим, оскільки значення функцій довіри і правдоподібності дорівнюють нулю (рис. 5 а) [7, с. 41].

2. Виділити чіткі зони, які підкреслюють зони розуцільнення, вузли перетину розломів різних напрямів, що є перспективними для наявності скупчень вуглеводнів, для північно-західного шельфу Чорного моря. Одночасно чітко визначились зони концентрації структур, виявлених геофізичними методами, хоча деякі виявлені і підготовлені до глибокого буріння структури потрапили в зони низької ймовірності (рис. 5 б) [3, с. 93].

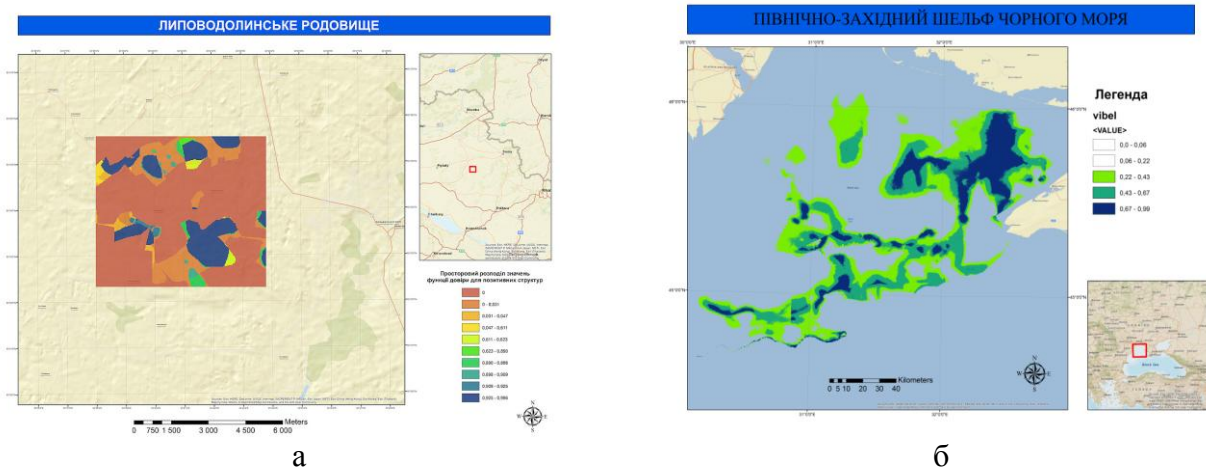


Рисунок 5 – Результат застосування комбінаційного правила Демпстера у задачі пошуку покладів нафти та газу: а) для Липоводолинського родовища [7]; б) для північно-західного шельфу Чорного моря [3]

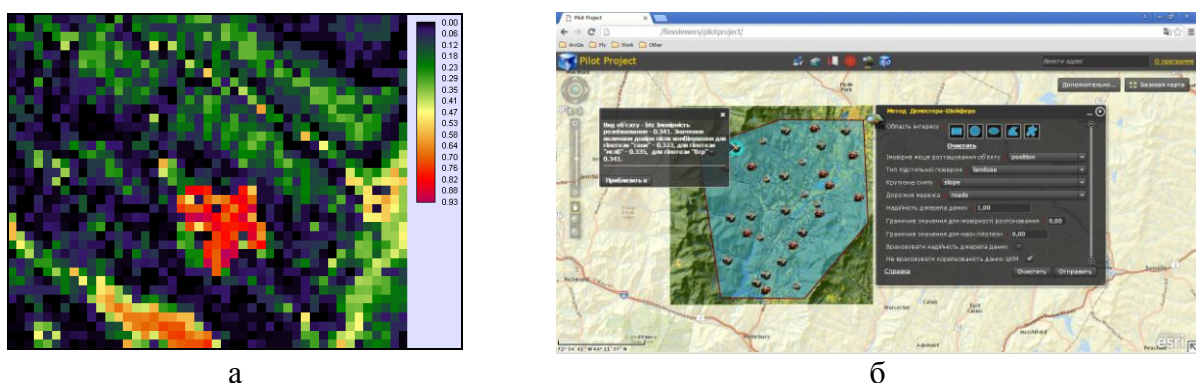


Рисунок 6 – Результат застосування комбінаційного правила Демпстера у задачі ідентифікації компактних штучних об'єктів: а) для багатоспектральних/гіперспектральних аерокосмічних зображень [9]; б) для даних аерокосмічного знімання та окремих шарів цифрової моделі місцевості

3. Виявити з достатньо високою точністю ($P \in [0,7; 0,94]$) на об'єднаному зображенні клас об'єктів, який не вдалось виділити на окремому супутниковому гіперспектральному зображенні (рис. 6 а) [9, с. 66].

4. Віднести кожний об'єкт інтересу до конкретного виду, підвищити достовірність його інтерпретації та аналізу в середньому на 44 % та зменшити рівень невизначеності на 13 % (рис. 6 б).

Проте, головним недоліком під час використання класичних положень ТСДШ для інтегрування гетерогенних ГПД на рівні рішень є спірність результату, що отримується, в умовах значного конфлікту між джерелами надходження інформації [2, с. 132], що підтверджується отриманими результатами.

Для нівелювання зазначеного недоліку авторами, керуючись положеннями, сформульованими вище у пп. 4–7, було розроблено метод, який дозволяє врахувати корельованість і узгодженість гіпотез та конфліктність джерел ГПД [3]. Його структурно-логічну схему наведено на рис. 7.

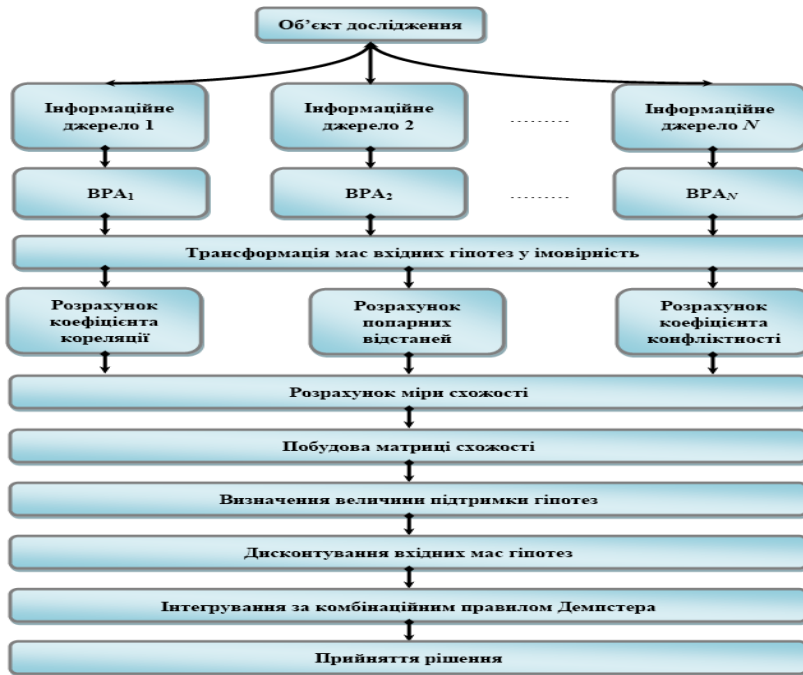


Рисунок 7 – Структурно-логічна схема запропонованого методу

Результати застосування розробленого методу для вирішення тих самих задач наведено на рис. 8.

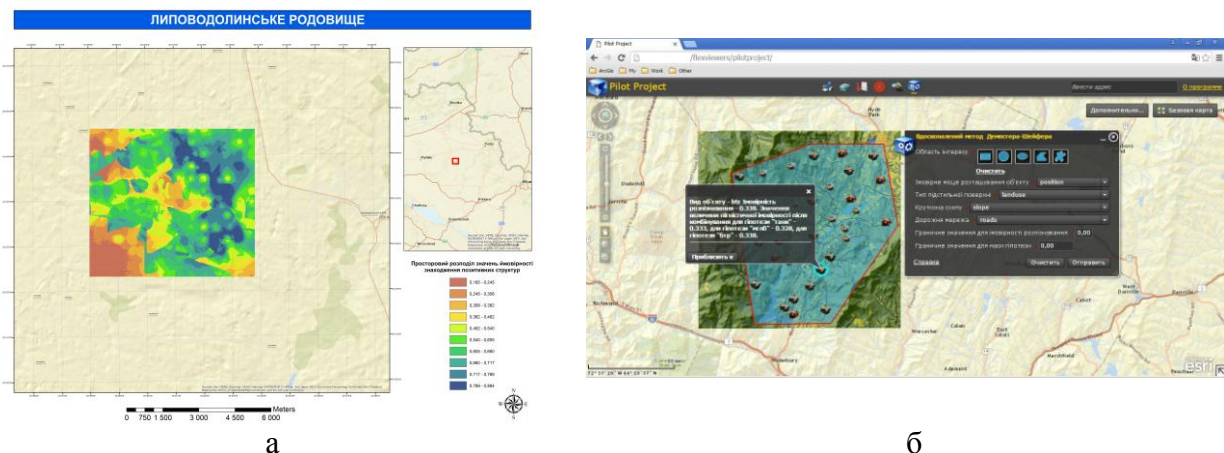


Рисунок 8 – Результат застосування запропонованого методу:

- а) у задачі пошуку покладів нафти та газу для Липоводолінського родовища;
- б) у задачі ідентифікації компактних штучних об'єктів для даних аерокосмічного знімання та окремих шарів цифрової моделі місцевості

1. Запропонований метод дозволяє віднести кожний об'єкт інтересу до конкретного виду, підвищити достовірність його інтерпретації та аналізу в середньому на 4 % та зменшити рівень невизначеності на 28 % (рис. 8 б).

2. Основна перевага запропонованого методу – позбавлення недоліків, що характерні для комбінаційного правила Демпстера, можливість переходу від інтервальної до точкової оцінки [7, с. 38].

9. Висновки

Запропоновано та обґрунтовано новий підхід до інтегрування гетерогенних геопросторових даних. Наведено структурно-логічну схему інтегрування, особливості його проведення на різних рівнях. Як науково-методичну платформу для проведення інтегрування обрано теорію свідчень Демпстера-Шейфера. Крім того, приведено загальні обмеження інтегрування, сформульовано нові підходи до оцінки ефективності виконання тематичних завдань та представлено приклади їх вирішення на рівні гіпотез (рішень).

Отримані результати дають підстави стверджувати про можливість застосування теорії свідчень Демпстера-Шейфера для проведення інтегрування гетерогенних геопросторових даних на рівні гіпотез (рішень) та її дієздатність у класичному вигляді лише для випадків, коли конфлікт між джерелами надходження інформації є мінімальним.

Для нівелювання зазначеного недоліку пропонується новий метод, який дозволяє врахувати корельованість і узгодженість гіпотез та конфліктність джерел геопросторових даних. Результати застосування запропонованого методу для тих же тематичних задач дозволяють зробити висновок, що він позбавлений недоліку, який характерний для комбінаційного правила Демпстера, і дає можливість перейти від інтервальної до точкової оцінки відповідної гіпотези.

Подальші дослідження мають бути спрямовані на узагальнення розвинених у статті методів інтегрування на випадки, коли необхідно здійснити комплексний аналіз та інтерпретацію суттєво неоднозначних даних (soft data), наприклад, експертних даних, і достатньо точних даних (hard data), наприклад, даних, отриманих технічними засобами видового спостереження.

СПИСОК ДЖЕРЕЛ

1. Le Hégat-Masclé S., Bloch I., Vidal-Madjar D. Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*. 1997. Vol. 35, N 4. P. 1018–1031.
2. Дюбуа Д., Прад А. Теория возможностей. Приложения к представлению знаний в информатике. Москва: Радио и связь, 1990. 228 с.
3. Novel Remote Sensing Methods for Mineral Prospecting: monograph / ed. V.I. Lyalko, M.O. Popov. K.: NASU, 2017. 221 p.
4. Josang A. Conditional reasoning with subjective logic. *Journal of Multiple-Valued Logic and Soft Computing*. 2008. Vol. 1, N 15. P. 5–38.
5. Fischer H.R. Abductive reasoning as a way of worldmaking. *The Impact of Radical Constructivism on Science*. 2001. Vol. 6, N 4. P. 361–383.
6. Heuer R. Psychology of intelligence analysis. Washington, DC: Center for Study of Intelligence. Central Intelligence Agency, 1999. 184 p.
7. Попов М.О., Станкевич С.А., Топольницький М.В., Титаренко О.В. Інтеграція дистанційних та геолого-геофізичних даних при пошуках покладів нафти та газу на суходолі. *Екологічна безпека та природокористування*. 2017. № 1–2 (23). С. 36–43.
8. Shafer G. A mathematical theory of evidence. Princeton: Princeton University Press, 1976. 297 p.
9. Попов М.О., Топольницький М.В. Классификация объектов на многоспектральных/гиперспектральных аэрокосмических изображениях на основе теории свидетельств Демпстера-Шейфера. *Математичні машини і системи*. 2014. № 1. С. 58–69.

10. Zeigler B.P., Praehofer H., Kim T.G. Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems. 2nd ed. NY: Academic Press, 2000. 510 p.
11. Chokr B., Kreinovich V. How far are we from complete knowledge? Complexity of knowledge acquisition in the Dempster-Shafer approach. *Advances in the Dempster-Shafer Theory of Evidence* / eds. R.R. Yager, J. Kacprzyk, M. Fedrizzi. NY: John Wiley & Sons, 1994. P. 555–576.
12. Kramosil I. Probabilistic Analysis of Belief Functions. NY: Kluwer Academic, 2001. 214 p.
13. Klir G.J., Wierman M.J. Uncertainty-Based Information: Elements of Generalized Information Theory. Berlin: Springer-Verlag Berlin Heidelberg, 1999. 168 p.
14. Yager R.R. On the Dempster–Shafer Framework and New Combination Rules. *Information Sciences*. 1987. Vol. 41. P. 93–137.
15. Dubois D., Prade H. On the combination of evidence in various mathematical frameworks. *Reliability Data Collection and Analysis*. 1992. P. 213–241.
16. Hall D.L., McMullen Sonya A.H. Mathematical techniques in multisensor data fusion. 2nd edition. Norwood, MA: Artech House, 2004. 449 p.
17. Богомолов Л.А. Дешифрирование аэроснимков. М.: Недра, 1976. 145 с.
18. Марр Дж. Информационный подход к изучению представления и обработки зрительных образов / пер. с англ. М.: Радио и связь, 1987. 400 с.
19. Аковецкий В.И. Дешифрирование снимков: учеб. для вузов. М.: Недра, 1983. 374 с.
20. Klir G.J. Uncertainty and information measures for imprecise probabilities: An overview. *Proc. of 1st International Symposium on Imprecise Probabilities and Their Applications* (Ghent, Belgium, 29 June–2 July 1999). Ghent, 1999. P. 234–240.
21. Sudano J. Pignistic probability transforms for mixes of low – and highprobability events. *Proc. of Int. Conf. on Information Fusion 2001* (Montreal, Canada, August 2001). Montreal, 2001. P. 23–27.

Стаття надійшла до редакції 22.05.2019