

ВИКОРИСТАННЯ ПРИХОВАНИХ МАРКОВСЬКИХ ПРОЦЕСІВ В ЗАДАЧАХ ПРОГНОЗУВАННЯ АНОМАЛЬНИХ ПРИРОДНИХ ЯВИЩ

Робота присвячена розробці математичного апарату для прогнозування аномальних природних явищ. Зроблено висновок щодо необхідності врахування екзогенних процесів в розрахунку показників екологічної складової сталого розвитку. Запропоновано підхід до ситуаційного аналізу екологічних процесів, який ґрунтується на використанні прихованих марківських моделей. Наведено перелік задач, які необхідно розв'язувати при використанні прихованих марковських моделей для прогнозування аномальних природних явищ. Перевагою даного підходу є те, що він дає можливість описувати як явно визначені процеси, так і приховані впливи, які, як правило, мають місце в екології.

Ключові слова: аномальні природні явища, прогнозування, приховані марковські моделі.

Вступ. Глобальність сучасних процесів, зростаючі обсяги інформації і час, що скорочується, на вироблення і прийняття управлінських рішень, вимагають створення інтелектуальних систем підтримки прийняття рішень, нагромадження й обробки інформаційної бази результатів контролю і безлічі типових рішень для осіб і організацій, що відповідальні за прийняття і виконання рішень в еколого-економічному середовищі. Рішення цієї задачі вимагає у свою чергу розробки нових підходів, моделей методів і алгоритмів з позиції досягнення єдиної мети в умовах різномірної невизначеності.

В останні роки стає зрозуміло, що основою розв'язання гострих соціально-економічних і ресурсо-екологічних проблем є перехід до моделі сталого розвитку. Концепція сталого розвитку економіки визнана світовою спільнотою народів домінантною ідеологією розвитку людської цивілізації у ХХІ ст., стратегічним напрямом забезпечення матеріального, соціального і духовного прогресу суспільства.

Сталий соціально-економічний розвиток означає, таке функціонування народногосподарського комплексу, коли одночасно забезпечуються: задоволення зростаючих матеріальних і духовних потреб населення; раціональне та екологічнобезпечне господарювання й високоефективне використання природних ресурсів; підтримання сприятливих для здоров'я людини природно-екологічних умов життєдіяльнос-

ті, збереження, відтворення і примноження якості довкілля та природно-ресурсного потенціалу суспільного виробництва. Інакше кажучи, сталий розвиток — це насамперед економічне зростання, за якого ефективно розв'язуються найважливіші проблеми життєзабезпечення суспільства без виснаження, деградації і забруднення довкілля.

У 20 столітті, особливо в останній його третині, у зв'язку з посиленням впливу людини на природу, екологія набула особливого значення як наукова основа раціонального природокористування і охорони живих організмів. З 80-х років 20ст. екологічні проблеми, породжені сучасним суспільним розвитком, викликали сплеск наукових досліджень, технічних розробок і законодавчих рішень, мета яких — охорона навколишнього середовища і стійкий економічний розвиток сучасного суспільства [1].

Але бурхливий ріст досліджень в області математичного моделювання в екології поки що не привів до створення узагальнюючих екологічних моделей, які погоджують всі елементи екологічних процесів у цілісну єдину систему.

Основна частина. Існують дві основні групи процесів по своєму генезису або природі походження:

– ендегенні (антропогенні), викликані господарською діяльністю людини, які можуть бути керованими й у визначених випадках здемпфованими, або навіть відверненими;

– екзогенні, що є проявом космічних і внутрішніх геологічних процесів і змін, які носять суцільно об'єктивний характер, не піддаються керуванню, але можуть спостерігатися, прогнозуватися, передбачатися й отже, можуть бути розроблені комплекси заходів для мінімізації можливого збитку.

В рамках теорії сталого розвитку обмежуються, здебільше, розглядом ендегенних процесів, але і екзогенні процеси так само впливають на навколишнє середовище і на якість людського життя

Екзогенні екологічно небезпечні процеси відрізняються складними взаємозв'язками, взаємозалежностями, взаємодіями різноманітних факторів і причин, мають наступні характерні властивості й особливості: різноманітність і різнотипність причин і факторів, дії яких приводять до їхнього виникнення; просторова розподіленість умов виникнення, невизначеність у часі і просторі динаміки розвитку і ре-

гійонів їхнього впливу на екосередовище; нестаціонарність властивостей і невизначеність їхніх характеристик.

Серед всієї кількості екзогенних процесів варто виділити клас процесів, які неможливо моделювати за допомогою динаміко-чисельних методів, а через відсутність певної періодичності (добової, місячної, річної, або іншої сталої періодичності) їх важко описувати за допомогою емпірико-статистичних методів. В природі саме такі процеси викликають найбільш сильний негативний вплив як на чинники, які впливають на якість життя людини, так і на саме природне середовище. Зазвичай це є шторми, тайфуни, цунамі, гірські обвали, каменепади, зсуви, селі, водна ерозія, паводки, зливи, снігопади, снігові лавини, землетруси. Такі процеси називають аномальними.

Аномальні природні процеси — це підмножина складних фізичних систем, із усіма їхніми властивостями — відкритістю, динамічністю, унікальністю, слабкою формалізуемістю, багатокритеріальністю в контексті вибору рішень, невизначеністю, що зумовлено неповнотою або відсутністю знань про природу даного процесу, обмеженою можливістю математичного опису й обчислювальної реалізації, складністю в застосуванні технічних засобів виміру або керування, наявністю стохастичного або суб'єктивного (волютаристичного) факторів.

Ціль роботи. Актуальною є задача розробки аналізу протікання таких процесів, а також розробити методи прогнозування таких процесів для подачі інформації в систему розрахунків відповідних індикаторів екологічної складової концепції сталого розвитку..

Методи вирішення задачі. Одним з можливих варіантів вирішення цієї задачі є підхід до ситуаційного аналізу екологічних процесів, який ґрунтуються на використанні прихованих марковських ланцюгів. Перевагою даного підходу є те, що він дає можливість описувати як явно визначені процеси, так і приховані впливи, які, як правило, мають місце в екології. Застосування такого підходу до ситуаційного аналізу в екології дасть можливість суттєво підвищити якість розпізнавання ситуацій та підвищити якість рішень щодо цих ситуацій. [2].

Різні розробки Марковських моделей на сьогодні мають широке застосування щодо вирівнювання послідовностей (білків, ДНК, тощо), а також для виявлення гомології між ними та пошуку й розпізнавання послідовностей, деяким узагальненою подібністю.

Під марковським ланцюжком в узагальненому розумінні маємо послідовність випадків, кожне з яких відбувається із певною ймовірністю. Перші застосування цього математичного апарату були для розпізнавання мови.

Головною задачею розпізнавання, яку ми досліджуємо, використовуючи математичний апарат марковських моделей, полягає в наступному. Нехай маємо дві довгих послідовності координат, потрібно виявити наскільки ці послідовності належать одній події, є близькими за своєю статистичною структурою. При цьому наш висновок повинен бути стійким щодо мало суттєвих змін, таких як вставки та випадання окремих дільниць послідовностей, переставляння невеликих фрагментів та ін.

Можна порівнювати й цілком невеликі послідовності, але тоді їх повинно бути достатньо. Статистична наближеність послідовностей може бути наслідком подоби їх функціональних властивостей або єдності виникнення (належить одному оператору). Наприклад, маючи набір послідовностей одного оператора, можна виявити чи належить цьому оператору деяка послідовність чи ні. Тип марковської моделі обирається в залежності від змісту завдання.

Перевагою даного підходу є те, що він дає можливість описувати як явно визначені процеси, так і приховані впливи, які, як правило, мають місце в екології. Застосування такого підходу до ситуаційного аналізу в екології дасть можливість суттєво підвищити якість розпізнавання ситуацій та підвищити якість рішень щодо цих ситуацій.

Проте для систем визначення параметрів процесу традиційна марковська модель не дає точного опису і може застосовуватися лише із значними обмеженнями, оскільки реальні системи визначення параметрів процесу мають структуру вкладених стохастичних систем.

Розглянутий вище тип марковських моделей може бути добре застосовний там, де стани моделі відповідають реально спостережуваним подіям, проте такі моделі дуже обмежені, щоб бути застосованими до великого числа різних завдань.

На практиці найбільш поширені так звані приховані марковські моделі (ПММ). Вони описують вкладені стохастичні процеси, коли реально спостерігаються тільки події зовнішнього процесу, а

події якогось прихованого процесу не спостерігаються безпосередньо – вони можуть бути визначені тільки із спостережень зовнішнього процесу. У таких моделях спостережувані події є імовірнісними функціями від реального стану системи.

Для повного визначення ПММ необхідно задати два параметри спостережуваних символів природного процесу (M і N), і три імовірнісні величини A, B і Π . Для компактності записують так: $\lambda = (A, B, \Pi)$, де

1) N – число станів екологічної моделі. Не дивлячись на те, що внутрішні стани системи приховані, для реальних додатків число станів N вибирається, виходячи з фізичних властивостей модельованого процесу. Позначатимемо конкретне значення стану як S і стан системи у момент часу t як S_t .

2) M – число різних, спостережуваних в кожному стані символів (тобто розмір алфавіту). Спостережувані символи відповідають фізичному виходу модельованої системи. Далі умовимося позначати конкретні символи як O_1, O_2, \dots, O_t

3) $A = \{a_{ij}\}$ – матриця вірогідності переходів, де $a_{ij} = P[q_{i+1} = S_j | q_i = S_i]$, $1 \leq i, j \leq N$

Для спеціального випадку, коли будь-який стан може бути доступний з будь-якого іншого за один крок (повнозв'язний граф), для всіх a_{ij} виконується умова $a_{ij} > 0$. Для інших випадків може бути рівною 0 для однієї або декількох пар i і j .

4) $B = \{b_j(k)\}$ – розподіл вірогідності спостережуваних символів в стані j

де $b_j(k) = P[v_k | q_i = S_j]$, $1 \leq j \leq N$, $1 \leq k \leq M$ (для безперервного випадку $b_j(k)$ задається як функція розподілу щільності вірогідності).

5) $\Pi = \{\Pi_i\}$ – вірогідність кожного початкового стану,

Основні завдання при застосуванні ПММ. Основні завдання при застосуванні ПММ до визначення параметрів процесів наступні. Для використання ПММ при розпізнаванні мови необхідно вирішити три задачі.

Завдання 1: Якщо задані послідовність спостережень природного процесу і екологічна модель $\lambda = (A, B, \Pi)$ то як ефективно обчислити $P(O | \lambda)$ - вірогідність такої послідовності при заданих параметрах моделі?

Завдання 2: Якщо задані послідовність спостережень природного процесу і екологічна модель $\lambda = (A, B, \Pi)$ то як визначити відповідну послідовність внутрішніх станів ?

Завдання 3: Якщо задані послідовність спостережень природного процесу, то як визначити параметри екологічної моделі $\lambda = (A, B, \Pi)$, виходячи з критерію максимізації $P(O | \lambda)$?

Всі три задачі мають свої алгоритми вирішення. [3].

Розв'язання першої задачі дозволить ефективно розпізнавати елементи з деякого обмеженого набору, маючи готові, вже навчені моделі цих елементів. В цьому випадку наявний елемент дискретизується і після сегментації на певні примітиви, розглядається як набір послідовностей спостережень. Замість безпосередньо графічного сигналу системи зазвичай працюють з якимсь набором примітивів (меншим за об'ємом), що характеризує наявний сигнал. Цей етап називається передобробкою сигналу. Потім кожна послідовність спостережень, що є невідомим словом, перевіряється на відповідність всім наявним в словнику моделям слів. Для цього обчислюється вірогідність генерації такої послідовності спостережень кожній з наявних в словнику моделей. Слово, модель якого з найбільшою вірогідністю генерує таку послідовність спостережень, і є результатом розпізнавання.

Розв'язання другої задачі дозволяє виявляти «приховану» частину моделі, тобто знаходити її внутрішні стани. Насправді визначити ту послідовність станів, яка мала місце, неможливо, а у випадку з визначанням такої послідовності просто немає, оскільки вважається, що модель мовоутворення людини не носить стохастичний характер. Проте, знаходження оптимальної послідовності дає нам додатковий критерій порівняння при пошуку оптимальної моделі і дозволяє виявляти статистичні характеристики конкретних станів моделі.

Розв'язання третього завдання дозволяє навчати моделі, тобто обчислювати параметри моделі так, щоб вона найкращим чином описувала так звані «тренувальні» послідовності. Завдання трену-

вання моделей вважається найскладнішим при використанні ПММ в системах визначення параметрів процесу, оскільки невідомо єдиного і універсального способу її рішення, а від результату навчання моделей залежить якість роботи системи розпізнавання.

Висновки. Оскільки традиційні ПММ погано моделюють тривалість стану (для цього доводиться послідовно включати в модель декілька однакових станів або підвищувати вірогідність переходу системи з даного стану в саме себе), то в них доводиться явно вводити щільність тривалості стану і забороняти переходи моделі з будь-якого стану в саме себе. Така ПММ носить назву «ПММ з тривалістю станів». На практиці в системах визначення параметрів це може бути корисно для того, щоб врахувати різницю в тривалості перебування сигналу в певному стані.

Таким чином за наявності певної статистики попередніх спостережень природних процесів можна з певною ймовірністю прогнозувати час появи, а також параметри протікання аномальних природних процесів різної природи.

ЛІТЕРАТУРА

1. Селін Ю.М. Системний аналіз екологічно небезпечних процесів різної природи / Ю.М. Селін // Системні дослідження та інформаційні технології. — 2007. — № 2. — С.22–32.
2. Baum L.E. «An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes», *Inequalities*, vol.3, pp. 1-8, 1972.
3. Rabiner L.R., «A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition», *Proceedings of the IEEE*, vol, 77. no.2, February 1989, pp. 257-284.