

УДК 336.1.07:351.72 + 004.93'1

Акімова О. В.

МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ ВИЯВЛЕННЯ СУМНІВНИХ ФІНАНСОВИХ ОПЕРАЦІЙ В СИСТЕМІ ФІНАНСОВОГО МОНІТОРИНГУ

Одним з заходів підвищення дієвості функціонування підсистеми фінансового моніторингу в системі державного фінансового контролю має стати удосконалення виявлення та ідентифікації ознак підозрілих операцій. Удосконалення нормативно-методичних документів з опису і деталізації ознак в її межах можна здійснювати декількома шляхами: 1) запровадження прецедентного підходу до ідентифікації правопорушень і злочинів, пов'язаних з легалізацією доходів, отриманих незаконним шляхом, та фінансуванням тероризму, в т.ч. запозиченням кейсів з закордонної практики, однак, формальність рішення або рекомендацій не означає їх правомірності, особливо за наявності розвиненої системи кодифікованого права. 2) деталізація фінансових та податкових питань адміністративного та кримінального законодавства, їх кодифікація й узагальнення, в т.ч. із залученням міжнародного досвіду; цей шлях також передбачає значну трудомісткість організації і здійснення фінансового моніторингу та впровадження процедур CDD/ECDD, KYC/KYCC, compliance у діяльність суб'єктів господарювання; 3) удосконалення методичного забезпечення виявлення сумнівних фінансових операцій.

У вузькому змісті результатом функціонування системи фінансового моніторингу є виявлення фактів позанормативної поведінки як таке. У зв'язку з цим варто відзначити, що результативність включає не тільки міру кількісного досягнення цілі, а й якісні характеристики: абсолютне охоплення, широта, інтенсивність та якість охоплення, ступінь точності ідентифікації випадків; швидкість процесу розслідування (від моменту отримання повідомлення до прийняття рішення про застосування санкцій); швидкість обробки взаємних запитів органів як здійснюють фінансовий моніторингу або протидію легалізації доходів; ступінь релевантності відповіді запиту та ін.

Ймовірність підвищення показників широти, інтенсивності та якості охоплення ФМ визначається ймовірністю отримання вірних рішень при розпізнаванні підозрілих фінансових операцій, які належать до різних типів позанормативної фінансової поведінки, яке в свою чергу залежить від помилок апріорної типологізації, визначення ознак кожного типу та їх порогових значень, розпізнавання об'єктів, а також кількості та властивостей ознак.

Ці завдання в більш інструментальному аспекті можна звести до двох:

1) як з масиву ознак, які вірогідно характеризують сумнівність характеру фінансової операції з позиції її зв'язаності з певним предикатним злочином виділити ті ознаки, які мають дискримінуюче значення для кожного типу предикатного злочину (позанормативної фінансової поведінки);

2) яким має бути конкретне чисельне значення характеристики для того, щоб на її основі можна було розмежувати нормативний та позанормативний характер здійснення фінансової операції для тих операцій, сумнівний зміст яких не визначається прямим зв'язком зі характером, метою, суб'єктом здійснення.

Розв'язання першої проблеми лежить в площині використання методів розпізнавання образів. Власне розпізнавання образів (*pattern recognition*) — це розділ теорії штучного інтелекту (*artificial intelligence*), що вивчає методи класифікації об'єктів. За традицією об'єкт, що піддається класифікації, називається образом (*pattern*). Образом може бути цифрова фотографія (розпізнавання зображень), буква або цифра (розпізнавання символів), запис мови (розпізнавання мови), наявність або відсутність у об'єкта певної характеристики тощо [1].

Ознаки образу – це якісні та кількісні параметри, сукупність яких дозволяє відрізнити один образ від іншого, робить образ індивідуальним. Ознаки можуть бути суттєвими (вони найбільшою мірою визначають індивідуальність, неповторність образу) і несуттєвими (вони меншою мірою впливають на визначення, розпізнавання і класифікацію образу), кількісними, якісними, апіорними, апостеріорними [2]. В такому разі розпізнавання образів (РО) – це процес розділу однотипних образів на класи за допомогою відповідних методів та засобів для виділення, ідентифікації, вимірювання та інтелектуальної обробки часто неповних, нечітких, непрямих, перекручених та суперечливих ознак образів.

Метою роботи є розробка методичного підходу до виявлення операцій, які підлягають фінансовому моніторингу за принципом розпізнавання образів.

Формально це представляє собою послідовність наступних дій:

1. Побудова типології позанормативної поведінки та сукупності ознак за кожним типом – створення масиву класів об'єктів A_k з атрибутами $\{a_{ij}\}$, які можуть бути мулевими змінними (наявність / відсутність певної ознаки) або чисельні значення з порогом або без порогу.

2. Кластеризація об'єктів A_k , для визначення схожості ознак позанормативності фінансової поведінки. Це дає змогу сформуванню структури об'єктів, щодо яких при проведенні фінансового моніторингу необхідно буде здійснювати диференційну діагностику, в т.ч. з використанням формалізованого (дискримінантний аналіз) або неформалізованого (експертна думка) методу.

3. Формування набору даних про фактичні значення ознак певного фактичного об'єкту $\Phi_m \{f_{ij}\}$ в ході ідентифікації підозрілості фінансової операції.

4. Порівняння фактичного масиву даних щодо ознак операції зі словником (базою даних) шаблонних ознак певного типу позанормативної фінансової поведінки.

При визначенні міри подібності об'єктів кластерного аналізу використовуються чотири види коефіцієнтів: коефіцієнти кореляції, показники віддалей, коефіцієнти асоціативності та ймовірності, коефіцієнти подібності. На практиці найбільшого розповсюдження у сфері соціальних та економічних наук здобули коефіцієнти кореляції та віддалей. В результаті аналізу сукупності вхідних даних створюються однорідні групи у такий спосіб, що об'єкти всередині цих груп подібні між собою за деяким критерієм, а об'єкти з різних груп відрізняються один від одного.

Кластеризація може здійснюватися двома основними способами, зокрема за допомогою ієрархічних чи ітераційних процедур.

Ієрархічні процедури – послідовні дії щодо формування кластерів різного рангу, підпорядкованих між собою за чітко встановленою ієрархією. Найчастіше ієрархічні процедури здійснюються шляхом агломеративних (об'єднувальних) дій. Вони передбачають такі операції:

послідовне об'єднання подібних об'єктів з утворенням матриці подібності об'єктів;

побудова дендрограми (деревоподібної діаграми), яка відображає послідовне об'єднання об'єктів у кластери;

формування із досліджуваної сукупності окремих кластерів на першому початковому етапі аналізу та об'єднання всіх об'єктів в одну велику групу на завершальному етапі аналізу.

Ітераційні процедури полягають в утворенні з первинних даних однорівневих (одного рангу) ієрархічно не підпорядкованих між собою кластерів.

Одним із найбільш поширених способів проведення ітераційних процедур ось уже понад сорок років виступає метод k-середніх. Його застосування потребує здійснення таких кроків:

розділення вихідних даних досліджуваної сукупності на задану кількість кластерів;

обчислення багатовимірних середніх (центрів тяжіння) виділених кластерів;

розрахунку Евклідової відстані кожної одиниці сукупності до визначених центрів тяжіння кластерів та побудова матриці відстаней, яка ґрунтується на метриці відстаней.

Використовують різні метрики відстаней, наприклад: Евклідова відстань (проста і зважена), Манхеттенська, Чебишева, Мінковського, Махаланобіса тощо.

Зміст принципу такого порівняння можна розуміти з прикладу розпізнавання мови або письмових знаків, коли якість розпізнавання пов'язується з наявністю різниці між розпізнаною та апріорною послідовностями. Для цього кожній парі послідовностей у відповідність ставиться певне число, яке відображає відстань між цими послідовностями. Таким числом може бути т.зв. відстань Левенштайна, яка дорівнює числу операцій вставки та заміни символів, необхідних для перетворення першої послідовності на другу [3].

Відстань Левенштайна обчислюється за допомогою алгоритму динамічного програмування, який мінімізує суму $S + D + I$, де S , D , I – число операцій вставки, видалення та заміни відповідно. Якщо позначити через H число збігів, а через N_1 та N_2 – потужність першої та другої послідовності, то виконуються тотожності:

$$N_1 = H + S + D; N_2 = H + S + I. \quad (1)$$

Відповідно середня точність розпізнавання послідовності визначається відносною кількістю помилково розпізнаних елементів послідовності:

$$EER = \frac{S+D+I}{N_1}, \quad (2)$$

де N_1 – кількість елементів апріорної (правильної) послідовності.

Іншим показником точності може бути середня кількість помилкових пар:

$$MER = \frac{S+D+I}{H+S+D+I} = 1 - \frac{H}{N}, \quad (3)$$

де $N = H + D + S + I$ – число пар ознак, які співставляються.

Окрім показника відстані мірою близькості фактичного образу до еталонного можуть бути показники спряженості [4]:

показник спряженості із підпростором, натягнутим на вектори ознак об'єктів з певного заданого типу:

$$R_i = \frac{x^T X_i [X_i^T X_i]^{-1} X_i^T x}{x^T x}, \quad (4)$$

де x – вектор ознак невідомого образу, який представлено для визначення відстані з i -м типом; X_i – матриця, складена з векторів образів еталонів, які належать до i -го типу операцій; показник спряженості з нуль-підпростором того ж простору:

$$S_i = \frac{x^T T_i T_i^T x}{x^T x}, \quad (5)$$

де T_i – матриця, складена із власних векторів, що відповідають нульовим власним значенням матриці $X_i^T X_i$.

Найбільш відомими та широко застосовуваними методами формування кластерів є: одиничного зв'язку; повного зв'язку; середнього зв'язку; метод Ворда:

метод одиничного зв'язку (метод близького сусіда) передбачає приєднання одиниці сукупності до кластера, якщо вона близька (знаходиться на одному рівні схожості) хоча б до одного представника цього кластера;

метод повного зв'язку (далекого сусіда) вимагає певного рівня подібності об'єкта (не менше граничного рівня), що передбачається включити у кластер, з будь-яким іншим;

метод середнього зв'язку ґрунтується на використанні середньої відстані між кандидатом на включення у кластер і представниками наявного кластера;

згідно методу Ворда приєднання об'єктів до кластерів здійснюється у випадку мінімального приросту внутрішньогрупової суми квадратів відхилень.

Оптимальною прийнято вважати кількість кластерів, яка визначається як різниця кількості спостережень і кількості кроків, після якої відстань об'єднання збільшується стрибкоподібно.

Апробацію можливостей кластерного аналізу можна зробити на прикладі побудови кластеру шаблонних ознак позанормативності окремих операцій в різних сферах фінансової діяльності, які дозволяють визначати ймовірну схожість випадків позанормативної поведінки за певними ознаками і звертати увагу на інші ознаки, які є дискримінуючими для ідентифікації конкретного типу позанормативної фінансової операції.

В табл. 1 наведено формування вихідних даних (у вигляді ймовірності сполучення типу поведінки і ознак сумнівності або позанормативності фінансових операцій) для проведення кластерного аналізу ознак за їх приналежністю до певних типів позанормативної фінансової поведінки за видом фінансової діяльності «фінансування зовнішньоекономічної діяльності».

Таблиця 1

Вихідні дані для проведення кластерного аналізу ознак позанормативної фінансової поведінки при фінансуванні ЗЕД

№	Тип поведінки	Ухилення від оподатк.	Шахрайство	Фінансування тероризму	Нераціональне використання державного майна
1	Належність до державного сектора	0,9	0,8	0,2	0,6
2	Різність місця реєстрації та рахунку	0,8	0,6	0,7	0,2
3	Використання ризикових юрисдикцій	0,8	0,9	0,5	0,1
4	Невідповідність вартості аналогічних контрактів	0,5	0,8	0,2	0,3
5	Різність митних вартостей	0,8	0,8	0,2	0,2
6	максимально скорочені терміни розрахунків	0,3	0,7	0,6	0,5
7	Повернення частини коштів як плата за послуги	0,1	0,9	0,6	0,8
8	Операції на допорогові суми	1	1	1	1
9	Авансовий платіж на суму ПДВ	0,9	0,8	0,1	0,3
10	Відсутність історії контрагента	0,5	0,9	0,9	0,2
11	Платежі фізичним особам	0,3	0,8	0,8	0,6
12	Предмет-нематеріальний актив	0,2	0,7	0,5	0,4
13	Різність місцезнаходження рахунку і реєстрації контрагента	0,4	0,8	0,8	0,3
14	Сплата без ввезення товару	0,2	0,7	0,4	0,5
15	Позастатутний характер діяльності	0,1	1	0,6	0,6

Результати цього аналізу показані на рис. 1. Вони свідчать, про те, що найменш важливою дискримінантною ознакою у визначенні типів позанормативної фінансової поведінки «ухилення від оподаткування», «шахрайство», «фінансування тероризму», «нераціональне використання державного майна» при фінансуванні ЗЕД. В той же час є два більш дрібних кластери, які поєднують ознаки, які типово супроводжують кожний з видів позанормативної фінансової поведінки: п перший (ознаки 6-7, 10-12, 14-15), та другий (ознаки 1-5, 9).

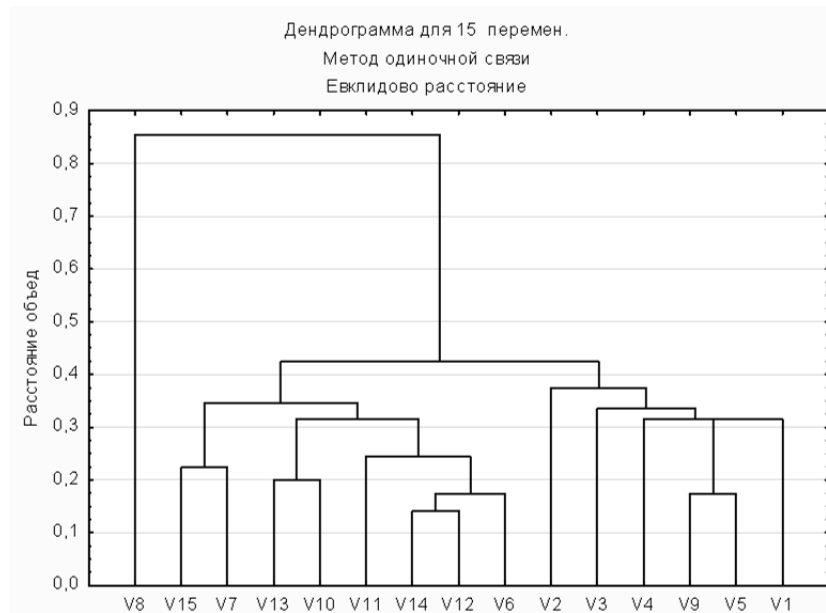


Рис. 1. Результати кластерного аналізу ознак окремих типів позанормативної фінансової поведінки для виду фінансової діяльності «Фінансування ЗЕД»

Дискримінантний аналіз – це інструмент, за допомогою якого за певним набором фактичних даних про фінансову операцію здійснюється оцінка подібності їх до типологічного об'єкту, визначається дискримінантна ознака, за якою можна судити про належність фактичного об'єкту до певного класу об'єктів з-поміж усіх.

Здійснення кластерного аналізу за оцінкою значущості певної ознаки (експертна оцінка ймовірності реалізації цієї ознаки за певного типу позанормативної фінансової поведінки) може дати картину схожості сумнівних ознак або забезпечити формування кластерів відповідності цих ознак певним типам позанормативної фінансової поведінки в рамках окремого виду фінансової діяльності.

Формування масиву даних про кількісні характеристики ознак окремих типів позанормативної поведінки, які є дискримінуючими в разі характерності ознаки і для нормативних і для позанормативних фінансових операцій передбачає побудова нечіткої моделі типу Мамдані з вербальним описом залежних залежної змінної.

Поняття «штучна нейронна мережа» виникло в галузі штучного інтелекту в процесі пошуку можливостей моделювання структури мозку, що надає змогу живим істотам навчатися, виправляючи власні помилки. Цим терміном позначають певний клас математичних моделей та їх програмних або апаратних реалізацій, які побудовані за принципами організації й функціонування біологічних нейронних мереж, тобто мереж нервових кліток живого організму. Особлива привабливість нейронних мереж полягає в тому, що їх застосування, завдяки так званому навчанню на наявних даних, дає можливість прогнозувати, які значення прийматимуть досліджувані змінні у нових спостереженнях, ґрунтуючись на даних попередніх спостережень. При грамотному застосуванні інструментарію нейронних мереж точність таких прогнозів значно перевищує точність прогнозів, здійснених за допомогою інших, зокрема класичних статистичних методів. Завдання прогнозування вирішується нейронними мережами аналогічно завданню розпізнавання образів, а умовою застосування нейромереж в прогнозуванні є наявність «історичних даних», використовуючи які мережа може «навчитися». Крім того, певні типи нейронних мереж (наприклад, мережі Кохонена) можуть застосовуватись й без навчаючих вибірок для вирішення завдань класифікації та кластеризації, що розширює можливості типологічного аналізу, оскільки з'являється новий інструмент перевірки якості типологій, які побудовані завдяки застосуванню інших методів [5].

Процес побудови нечіткої моделі відбувається в декілька етапів: визначення межі зміни вхідних параметрів (тобто максимальної та мінімальної оцінки певної якісної характеристики експертним шляхом або методом аналізу ієрархій); визначення методу побудови функцій належності вхідних змінних і досліджуваного показника у вигляді нечітких множин; створення бази знань; формалізація бази знань у вигляді нечітких логічних рівнянь; тестування нечіткої моделі; навчання нечіткої моделі.

Постановка задачі в даному випадку має наступний вигляд: обираються показники або параметри, які характеризують певний стан системи розпізнавання за певним напрямом (складовою результативності) $X = \{x_1, x_2, \dots, x_n\}$ та певний результуючий показник Y (він може являти собою або інтегральний показник, побудований згоранням відносно об'єктивних показників, індекс розвитку за певним напрямом тощо), та визначається вага g_i кожного показника множини X . Для якісних характеристик, які отримані як результати експертних оцінок, соціальних опитувань доцільно скористуватися методом побудови функцій належності на основі експертних оцінок.

Враховуючи те, що інтервальне значення кожної лінгвістичної змінної визначена експертами, тому доцільне використання трапецієвидної функції належності, верхня основа трапеції якої відповідає повній впевненості експерта в правильності своєї класифікації, а нижня – впевненості в тому, що жодні інші значення інтервалу $[0, 1]$ не потрапляють у обрану нечітку підмножину.

Оцінку навчання нейромережної моделі варто здійснювати з використанням показника середньої похибки, а тестування її чутливості та специфічності – з використанням аналізу ROC-кривих для визначення порогових значень ознак за окремими a_{ij} для формування еталону підозрілості фінансових операцій.

ВИСНОВКИ

На основі доведення положення про те, що ймовірність підвищення показників широти, інтенсивності та якості охоплення ФМ визначається ймовірністю отримання вірних рішень при розпізнаванні підозрілих фінансових операцій, запропоновано методичний підхід до виявлення операцій, які підлягають ФМ, що базується на побудові типології позанормативної поведінки та сукупності ознак за кожним типом, проведенні кластерного та дискримінаційного аналізу як основи диференціальної діагностики окремих типів позанормативної фінансової поведінки, створенні нечітких моделей типу Мамдані з вербальним описом залежних змінних, тестування їх чутливості та специфічності з використанням аналізу ROC-кривих для визначення порогових значень за окремими ознаками.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Лекції з курсу «Прикладні проблеми теорії розпізнавання образів» [Електронний ресурс]. – Режим доступу: <http://om.univ.kiev.ua/ua/page/courses>.
2. Кутковецький В. Я. Розпізнавання образів : навч. посіб. / В. Я. Кутковецький ; Миколаїв. держ. гуманіт. ун-т ім. П. Могили комплексу "Києво-Могилян. акад.". – Миколаїв, 2003. – 196 с.
3. Саакян А. А. Исследование свойств показателей качества систем распознавания речи / А. А. Саакян // Проблемы управления. – 2009. – № 4. – С. 66–73.
4. Козин Н. Е. Распознавание лиц по показателям сопряженности в пространстве суммирующих инвариантов / Н. Е. Козин, В. А. Фурсов // Компьютерная оптика. – 2008. – Т. 32, №4. – С. 400–402.
5. Кислова О. М. Можливості застосування штучних нейронних мереж в аналізі соціологічної інформації / О. М. Кислова, К. Б. Бондаренко // Вісник Харківського національного університету ім. В. Н. Каразіна : Соціологічні дослідження сучасного суспільства : методологія, теорія, методи. – 2010. – № 891. – Вип. 26. – С. 78–82.

Стаття надійшла до редакції 10.06.2016 р.