

ЗАСТОСУВАННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДО РОЗВ'ЯЗАННЯ ЗАДАЧ АКТУАРНОГО МОДЕЛЮВАННЯ ТА ОЦІНЮВАННЯ ФІНАНСОВИХ РИЗИКІВ

С.В. ДУБІНІНА, П.І. БІДЮК

Анотація. Подано результати застосування методів інтелектуального аналізу даних до актуарного моделювання та оцінювання ризиків страхових компаній. Як математичний апарат використано узагальнені лінійні моделі, метод групового врахування аргументів та нечіткий метод групового врахування аргументів, а також байєсівський підхід до оцінювання невідомих параметрів моделей. На підставі фактичних статистичних даних з галузі страхування побудовано нові узагальнені лінійні моделі для подальшого застосування під час оцінювання ризику втрат страхових компаній, розроблено й апробовано мережу Байєса для оцінювання ризику банкрутства страхових компаній у разі настання страхового випадку. Прийнятною для подальшого використання виявилась модель з гамма-розподілом та логарифмічною функцією зв'язку, параметри якої отримано за чотири ітерації алгоритму оцінювання. Великий ризик банкрутства страхових компаній свідчить про відсутність ефективного механізму управління коштами як власного капіталу, так і надходжень, отриманих від страхових договорів. Застосування методів інтелектуального аналізу даних є ефективним підходом до розв'язання задач прогнозування та оцінювання ризиків актуарних процесів.

Ключові слова: актуарні процеси, страхові ризики, узагальнені лінійні моделі, байєсівський підхід, методи інтелектуального аналізу даних, метод групового врахування аргументів, нечіткий метод групового врахування аргументів.

ВСТУП

У зв'язку з удосконаленням технологій пошуку та оброблення даних у суспільстві нагромаджуються великі масиви даних різного спрямування. Функціонування будь-яких підприємств супроводжується реєстрацією і записом детальної інформації щодо їх діяльності та споживачів. У результаті величезні скупчення даних утворюють хаотичні нагромадження, які без спеціального оброблення непотрібні; більше того, вони суттєво ускладнюють процес пошуку потрібної інформації. Критеріями ефективного оброблення даних є врахування таких особливостей: дані мають необмежений обсяг і є різнорідними; результати мають бути конкретними і зрозумілими, а інструменти для їх оброблення — простими та зручними у використанні.

Одним з методів ефективного оброблення інформації є інтелектуальний аналіз даних (ІАД), який полягає у виявленні у первинних даних, раніше невідомих, доступних, практично корисних і нетривіальних інтерпретацій знань, необхідних для прийняття рішень у різних галузях діяльності. До методів і алгоритмів ІАД належать: нейронні мережі, дерева рішень, символічні правила, методи найближчого сусіда і k -найближчого сусіда, метод опорних векторів, байєсівські мережі (БМ), лінійна регресія, кореляційно-регресійний

аналіз, ієрархічні методи кластерного аналізу, неієрархічні методи кластерного аналізу, зокрема алгоритми k -середніх і k -медіани; методи пошуку асоціативних правил, зокрема алгоритм Аргіогі; метод обмеженого перебирання; еволюційне програмування і генетичні алгоритми; різноманітні методи візуалізації даних і т. ін. [1–4]. Інтелектуальний аналіз даних виник і набув подальшого розвитку на базі досягнень прикладної статистики, розпізнавання образів, методів штучного інтелекту, теорії баз даних тощо. Одним з актуальних застосувань ІАД є аналіз даних у страховій діяльності.

Страхова (актуарна) діяльність спрямована на перерозподіл грошових коштів та акумулювання їх безпосередньо для страхової діяльності; інвестування цих коштів у різні галузі діяльності сприяє їх подальшому розвитку. Таким чином, виникають задачі аналізу фінансових процесів та менеджменту фінансових ризиків з використанням сучасних методів математичного моделювання, прогнозування та ефективної підтримки прийняття рішень із застосуванням методів ІАД.

Актуальність дослідження полягає у застосуванні методів ІАД до моделювання актуарних процесів та оцінювання фінансових ризиків страхових компаній (СК). Це пояснюється тим, що сьогодні сучасний стан фінансів СК потребує пошуку нових форм та методів підвищення їх конкурентоспроможності, удосконалення та впровадження у практику економіко-математичних моделей для оцінювання і прогнозування їх подальшого розвитку в умовах невизначеності, ризику реалізації багатьох процесів, що відбуваються у діяльності фізичних осіб та підприємств різних форм власності. Розмаїття форм прояву ризику, частота та складність наслідків його реалізації зумовлюють необхідність поглибленого аналізу ризиків та економіко-математичного обґрунтування фінансової політики СК[5]. Використання економіко-математичних методів для аналізу та управління фінансовими ризиками зумовлено їх можливістю отримання обґрунтованої та достовірної оцінки основоположних характеристик фінансової стабільності. До основних характеристик фінансової стабільності СК належать такі показники: імовірність розорення, маржа платоспроможності, власний капітал, страхові тарифи. Отже, задача створення систем для ефективного оцінювання фінансового стану з метою уникнення і недопущення банкрутства СК та підвищення рівня її фінансової стабільності є надзвичайно актуальним завданням.

ПОСТАНОВКА ЗАДАЧІ

Мета роботи — застосування методів ІАД для моделювання актуарних процесів і оцінювання фінансових ризиків страхових компаній. Для досягнення поставленої мети необхідно: дослідити задачу оцінювання фінансових ризиків у страхуванні з використанням методів ІАД; на основі запропонованої методики моделювання з використанням структур узагальнених лінійних моделей (УЛМ) побудувати прогнозні моделі для оцінювання ризику втрат СК у разі настання страхового випадку; порівняти якість отриманих результатів за методом групового врахування аргументів (МГВА) та нечітким МГВА; розробити БМ для оцінювання ризику банкрутства СК; сформулювати аналіз отриманих результатів та зробити висновки щодо доцільності застосування ІАД в актуарному моделюванні та прогнозуванні.

МОЖЛИВОСТІ ЗАСТОСУВАННЯ ІАД У МОДЕЛЮВАННІ СТРАХОВИХ РИЗИКІВ

Широкого застосування ІАД набув у актуарних задачах, оскільки візуальні інструменти дають змогу виконувати аналіз даних предметними фахівцями, які не мають відповідних математичних знань. Інтелектуальний аналіз даних має велику значущість для керівників і актуаріїв у сфері страхування. За допомогою методів ІАД можна домогтися відчутних переваг у конкурентній боротьбі за допомогою прогнозування страхових виплат у разі настання страхового випадку; оцінювання фінансових ризиків СК для недопущення подальшого банкрутства; виявлення шахрайства та ін. Досвід багатьох підприємств свідчить про доцільність застосування ІАД.

Ризики — можлива небезпека втрат, зумовлена певними природними явищами (землетруси, повені, засуха тощо), випадковими подіями техніко-технологічного характеру (пожежа, аварії, вибухи тощо), кримінальними явищами (крадіжки, розбійні напади, шахрайство тощо) та явищами суто економічного характеру, розмаїтість яких стрімко зростає у зв'язку з поглибленням суспільного поділу праці, розвитком кредитної системи (порушення ритмічності поставок, криза неплатежів, інфляція тощо).

Страхова діяльність супроводжується страховим ризиком. Але головним фундаментом страхової справи є розподіл ризику. Ризик є об'єктивною передумовою виникнення страхових відносин: немає ризику — немає потреби в страхуванні. Однак не кожен ризик може лягти в основу страхових відносин. Застрахованим буде лише той ризик, стосовно якого можна оцінити ймовірність настання страхового випадку, визначити розмір можливого збитку та нарахувати еквівалентну страхову премію. Під страховим ризиком розуміють прогнозний збиток об'єкта страхування в результаті настання страхової події [6]. Розв'язання задачі оцінювання фінансових ризиків у сфері страхування є одним з найважливіших етапів фінансового аналізу, оскільки для управління ризиком його потрібно спочатку вміти проаналізувати, а потім оцінити. Під оцінюванням ризиків розуміють систематичний процес виявлення факторів і видів ризиків, їх кількісну оцінку, тобто методологія аналізу ризиків поєднує взаємодоповняльні кількісний та якісний підходи.

Теоретичні дослідження проблеми невизначеності здебільшого присвячені аналізу причин виникнення, класифікації та методам кількісної оцінки ризиків, проте менше уваги приділяється якісним методам з урахуванням впливу неекономічних факторів на фінансово-господарську діяльність підприємств, які для практичного застосування були б адекватнішими до реальних можливостей вітчизняних підприємств, а отримані результати мали б підґрунтя для застосування у подальшому аналізі. У межах цієї роботи виконується дослідження кількісних та якісних методів аналізу фінансових ризиків з метою впровадження їх при оцінюванні і розрахунку міри впливу індикаторів економічного та неекономічного характеру.

Оскільки пошук імовірності розорення СК є однією із найважливіших задач страхової математики, на основі якої будуються основні актуарні концепції щодо фінансової оцінки стабільності, — врахування не тільки відсутності банкрутства, а і його недопущення, то зростає потреба у застосуванні

економіко-математичних методів до моделювання та аналізу фінансових ризиків і побудови актуарних моделей для пошуку оптимальної (справедливої, раціональної) величини страхової премії на основі знайденої імовірності розорення. Можливість побудови і застосування актуарних моделей пояснюється можливістю робити припущення щодо розподілу розмірів страхових виплат та інтервалів часу між виплатами, покладеними в основу моделі. Виплати можуть мати однаковий розподіл з відомою функцією розподілу або з довільною функцією розподілу; інтервали між виплатами можуть мати неоднакові параметри розподілу, а послідовність виплат описуватися за допомогою пуассонівського процесу з урахуванням додаткових можливостей виплати дивідендів учасникам.

МЕТОДИ ОЦІНЮВАННЯ ФІНАНСОВИХ РИЗИКІВ У СТРАХУВАННІ

Ступінь (міра) ризику — це ймовірність настання випадку втрат, а також розмір можливих збитків від нього. Імовірність настання такої події може бути визначена як об'єктивними, так і суб'єктивними методами. Об'єктивний метод визначення ймовірності настання події ґрунтується на обчисленні частоти, з якою відбувається ця подія, а суб'єктивний метод — на використанні суб'єктивних критеріїв, що ґрунтуються на різних припущеннях. Отже, задача оцінювання фінансових ризиків полягає у знаходженні залежності між визначеними розмірами втрат і ймовірністю їх виникнення. Сьогодні існують ефективні математичні моделі для оцінювання фінансових ризиків у страхуванні. Наприклад, відомі стандарти Basel II і Solvency I стимулюють розвиток моделей для оцінювання та менеджменту ризиків у фінансовій сфері.

Для оцінювання ризику в страховій практиці використовують спеціальні методи, які постійно розвиваються і удосконалюються. Так, на практиці є два підходи до аналізу ризиків — з власне факторів ризику і з позиції наслідків дії ризикових подій. Виходячи з цього моделі для оцінювання ризиків поділяють на два класи: моделі, засновані на аналізі наслідків (або «високорівневі моделі»), та моделі, засновані на аналізі факторів ризику (або «низькорівневі моделі»). Як ознаку для класифікації моделей використовують множину завдань, для розв'язання яких цю модель можна використовувати. Наприклад, моделі, які дають можливість отримати у явному вигляді функцію розподілу втрат, використовують для оцінювання середніх втрат СК і максимальних втрат на заданому рівні значущості (табл. 1).

Таблиця 1. Класифікація моделей для аналізу і менеджменту фінансових ризиків

Клас моделей	Застосування для розподілу ризиків за групами критичності	Застосування для визначення середніх втрат	Застосування для визначення максимально можливих втрат	Застосування для виявлення подій, що ведуть до настання ризику
Моделі, засновані на аналізі наслідків (або високорівневі моделі)				
BIA	–	–	+	–
LDA	+	+	+	–
IMA	–	–	+	–

Моделі, засновані на факторах ризику (або низькорівневі моделі)				
Sb-АМА	+	+	+	–
Метод функціональних кореляцій	+	+	+	–
Регресійний аналіз	+	+	+	+
Байєсівські мережі	+	+	+	+
Методи нечіткої логіки	+	–	–	+

Моделі, засновані на аналізі наслідків. Цей клас методів ґрунтується на апараті математичної статистики. Більшість таких методів припускає збирання і дослідження даних про втрати, зумовлені фінансовим ризиком упродовж попередніх періодів з подальшою екстраполяцією втрат на наступні періоди. Найбільш поширеними є такі моделі [5–8]:

BIA (Basic Indicator Approach — метод базових показників). Модель ВІА описує вимоги до забезпечення достатності капіталу для покриття ризиків. Ця модель запропонована для оцінювання максимально можливих втрат від операційного ризику на заданому рівні значущості (наприклад, 99,0 %). Вона ґрунтується на припущенні про те, що операційні втрати є залишковими від загальних втрат, спричинених ризиками. Квантиль розподілу випадкової величини X , що характеризує величину втрат, задається формулою

$$\hat{Q}_{99\%} = \bar{G}_I \alpha,$$

де $\hat{Q}_{99\%}$ — оцінка 99%-го квантиля випадкової величини, що характеризує втрати; \bar{G}_I — середній валовий дохід за три попередні роки, протягом яких валовий річний дохід був додатним; $\alpha=15\%$ — коефіцієнт, визначений на основі дослідження галузевої практики (для фінансових організацій, банків) Базельським комітетом. Значення цього коефіцієнта зазвичай уточнюють з використанням результатів оцінювання втрат, спричинених фінансовим ризиком в українських СК. Недоліком моделі ВІА є те, що отримана оцінка обсягу можливих фінансових втрат залежатиме тільки від обсягу бізнесу і не залежатиме від особливостей страхового портфеля.

LDA (Loss Distribution Approach — метод розподілу збитків). В основі моделей класу LDA лежить припущення про те, що випадкову величину X , яка характеризує розмір втрат, що сталися протягом часу t , можна визначити так:

$$x = \sum_{i=1}^{n(t)} L_i,$$

де $n(t)$ — випадкова величина, яка характеризує кількість випадків утрат конкретного типу за певний період часу t ; L_i — множина випадкових величин, що характеризують величини можливих окремих втрат. Припуска-

ється, що величини з множини L_i — незалежні й однаково розподілені для конкретного типу збитків. Для побудови таких моделей виконується аналіз втрат за вибраний період для кожної пари «лінія бізнесу»/«тип втрат». Для кожної такої пари на підставі даних про частоту втрат і величину збитків, що спостерігались у минулому, розраховується вибіркоче середнє значення частоти виникнення ризикових подій $E(n(t))$ і вибіркоче середнє значення величини втрат у разі виникнення ризикової події $E(L)$. Після визначення функцій розподілу випадкових величин $n(t)$ і L будується функція розподілу випадкової величини x (зазвичай за методом Монте-Карло) [8]. Функція розподілу випадкової величини x , що характеризує загальний обсяг втрат за вибраним видом ризиків, дає змогу знайти точкову оцінку математичного сподівання втрат і розрахувати квантиль заданого рівня, тобто значення OpVaR (Operational Value-at-Risk).

Варіанти методу LDA поділяють на дві групи. Перша група моделей ґрунтується на безпосередньому аналізі ризикових подій і зумовлених ними втрат без урахування факторів ризику та причинно-наслідкових зв'язків у межах складнішої моделі. До таких моделей можна віднести як безпосередню оцінку характеристик розподілу частоти і дії ризикових подій, так і методики, що спираються на теорію екстремальних значень [5, 7]. Друга група моделей ґрунтується на розширеній множині змінних, до якої входять і фактори ризику.

ІМА (Internal Measurements Approach — метод внутрішніх вимірів). Метод ІМА дає змогу обчислити оцінку максимально можливих втрат для конкретного виду ризиків без побудови розподілу випадкової величини x , що характеризує розмір утрат. В основі цього підходу лежить таке припущення: якщо поділити всі збитки на очікувані (тобто близькі до математичного сподівання суми збитків за період) і непередбачені (такі, що перевищують середнє значення і належать до «хвоста» статистичного розподілу), то існує функціональна залежність між величинами очікуваних і непередбачених збитків. Простим випадком є лінійна залежність, тобто

$$\hat{Q}_{99\%} = E_1 P_E L_{GE} = \gamma E_L \alpha,$$

де $\hat{Q}_{99\%}$ — оцінка 99%-го квантиля розподілу можливих втрат (величини капіталу, необхідного для покриття ризику) за цим типом події; P_E — імовірність настання несприятливої події цього типу в даний період; L_{GE} — середня величина збитку конкретного типу за умови настання несприятливої події; E_L — коефіцієнт масштабу; γ — коефіцієнт для оцінювання вимог до капіталу через оцінку очікуваних збитків E_L . Сукупна оцінка максимально можливих втрат визначається сумою усіх оцінок, розрахованих для різних типів ризикових подій і ліній бізнесу.

Статистичний метод оцінювання ризику. Цей метод полягає у вивченні статистики втрат (негативних наслідків реалізації рішень) в аналогічних видах підприємницької діяльності. При цьому можуть використовуватись різні способи оцінювання, у тому числі і дисперсійний аналіз. Основним показником, який розраховується на підставі статистичного методу, є частота втрат, пов'язаних із певним видом діяльності:

$$f = \frac{n'}{n_{\text{заг}}},$$

де n' — кількість випадків настання втрат у статистичній вибірці; $n_{\text{заг}}$ — загальна кількість випадків, що розглядались у статистичній вибірці. При прийнятті рішення на основі цього методу показник частоти втрат переноситься на прогнозні дані і розглядається вже як імовірність настання втрат певного рівня. Натепер статистичний метод застосовують у різних модифікаціях і найбільш поширеним стає метод статистичного моделювання (метод Монте-Карло). Перевагою цього методу є можливість аналізувати й оцінювати різні сценарії розвитку досліджуваного проекту з урахуванням різних факторів впливу в межах одного підходу. Недолік методу — великий рівень використання імовірнісних характеристик, що іноді не задовольняє менеджерів проекту.

Середні величини дають можливість виразити типові розміри ознак якісно однорідних явищ та виміряти їх коливання в околі середнього рівня розвитку. Метод середніх величин є одним зі статистичних методів дослідження. У процесі оцінювання ризику передбачається розмежування окремих ризикових груп на дрібніші підгрупи з метою створення аналітичної бази для визначення ризику за певними ризиковими ознаками. Метод процентів належить до групи методів статистичного аналізу і у термінах оцінювання ризиків означає деяку сукупність додатних і від'ємних відхилень від середнього ризикового типу наявної аналітичної бази. Такі відхилення виражають у процентах або в проміле від середнього ризикового типу.

Використовуються також економетричні, статистичні методи оцінювання й аналізу ризиків, методи вербального аналізу тощо. У сучасних умовах вітчизняні та зарубіжні вчені володіють потужним аналітичним інструментарієм для оцінювання та відстежування ризику; застосовують актуарні розрахунки, моніторинг, комплексне моделювання страхових процесів, емпіричний досвід, методи експертних оцінок, асоціацій та аналогій, експертиз тощо.

Моделі на основі аналізу факторів ризику передбачають поглиблений аналіз процесів у організації і дають змогу використовувати інформацію про внутрішні причинно-наслідкові зв'язки. У межах моделей цього класу використовують різні математичні методи, зокрема Sb-АМА (Scenario-based Advanced Measurement Approach — сценарний аналіз); метод функціональних кореляцій; регресійні моделі; байєсівські мережі; методи нечіткої логіки та деякі інші.

Метод Sb-АМА ґрунтується на визначенні факторів ризику (тобто можливих джерел ризиків), на основі яких генеруються сценарії за правилом «що буде, якщо». Таким чином, на відміну від описаних вище методів, які припускають аналіз збитків, що сталися, ця модель ґрунтується на оцінюванні втрат, які можливі в майбутньому при реалізації деякої події. Для кожного сценарію на підставі історичних даних або експертних оцінок визначають частоту і розмір збитків. Після перевірки даних і виправлення некоректних оцінок сценарії групують за факторами ризику і для кожної групи сценаріїв оцінюють параметри статистичних розподілів частоти і величини збитків. Завершальним етапом є імітаційне моделювання за методом Монте-Карло, що дозволяє оцінити загальний розподіл збитків [9, 10]. Оцінювання виконується за припущення, що відома сім'я, до якої нале-

жить розподіл випадкових величин кількості збитків (наприклад, розподіл Пуассона) і величини збитків (розподіли з «важкими хвостами», наприклад лог-нормальний або розподіл Парето). Отже, постійний контроль за фінансовим станом у такий спосіб дає змогу заздалегідь прослідкувати негативні тенденції і тим самим зменшити ймовірність ризику банкрутства підприємства.

Метод функціональних кореляцій ґрунтується на створенні структурної моделі організації. У цьому випадку перехід до математичної моделі здійснюється через побудову орієнтованого графу, вершини якого відповідають співробітникам і підрозділам, а ребра — потокам інформації. Кожній вершині ставиться у відповідність випадковий процес, що відображає її функціональність. Особливістю методу є встановлення стохастичних залежностей між функціональністю зв'язаних вершин. При цьому аналіз ризиків виконується не стосовно наслідків або окремих факторів ризику, а стосовно співробітників, систем, бізнес-процесів, яким поставлено у відповідність вершини графу.

Аналогічно до методу LDA вводиться розподіл величин втрат (для вершини i цей розподіл реалізується випадковою величиною L_i), пов'язаних з окремими випадками реалізації ризиків. Випадки реалізації ризиків описуються для вершини i випадковим процесом $n_i(t)$, який має два стани: $n_i(t) = 0$, що відповідає нормальному функціонуванню, і $n_i(t) = 1$, що означає збій. Сукупні втрати $X_i(t + \Delta t)$ на момент $(t + \Delta t)$, спричинені збоями на i -й вершині графу, можна задати так:

$$X_i(t + \Delta t) = X_i(t) + n_i(t + \Delta t)L_{i,t+\Delta t},$$

де $L_{i,t+\Delta t}$ — реалізація випадкової величини L_i . Для задання зв'язку між різними функціональними елементами вводиться поняття «підтримка», яку отримує кожний функціональний елемент від пов'язаних з ним елементів. У випадку, якщо обсяг підтримки стає меншим від заданої величини, відповідний функціональний елемент переходить у стан відмови (тобто для поставленої йому у відповідність i -ї вершини графу покладається $n_i(t) = 0$). Для відображення цієї структури в модель уводяться випадкові процеси $h_i(t)$:

$$h_i(t) = \vartheta_i - \sum_j w_{ij}n_j(t) + \varepsilon_i(t),$$

де ϑ_i — середній обсяг підтримки, яку i -й процес отримує у повністю функціональній системі (тобто, коли всі $n_i(t) = 1$); w_{ij} — обсяг підтримки, яку i -й процес отримує від j -го процесу (і відповідно втрачає, коли j -й процес перебуває у стані відмови); $\varepsilon_i(t)$ — гаусів білий шум. Цей метод орієнтовано на аналіз функціональності процесів і виявлення збоїв. Фінансові ризики мають іншу природу, пов'язану з відсутністю коштів на відшкодування страхового випадку, який настав за відповідним страховим договором.

Регресійні моделі ґрунтуються на виявленні причинно-наслідкових зв'язків між спостережуваними індикаторами і рівнем ризику. Розрізняють дві основні групи показників, які можна використовувати як спостережувані індикатори (пояснювальні змінні): 1) змінні оточення — це кількісні показ-

ники, що характеризують бізнес-процеси підприємства; 2) фактори ризику — кількісні показники, що характеризують спостережувані випадки реалізації ризиків. Така математична модель має вигляд

$$x = Af + b + \varepsilon,$$

де x — величина втрат, зумовлених фінансовим ризиком; f — вектор значень спостережуваних змінних; ε — випадкова величина, що визначає похибку моделі; A і b — оцінювані параметри, що характеризують залежність між змінною x та факторними змінними f . Для застосування цього методу необхідно мати достатній обсяг даних, щоб забезпечити високу точність оцінок.

Методи нечіткої логіки дають можливість найбільш повно використувати експертне оцінювання для аналізу ризиків у тих випадках, коли точних даних немає або вони неповні. Нечітка логіка наближує модель до міркувань людини в процесі прийняття і обґрунтування рішень. Методи нечіткої логіки можна застосовувати для оцінювання обсягу втрат і виявлення ризику банкрутства СК. На зарубіжних ринках часто застосовують автоматизовані системи для оцінювання фінансових ризиків, засновані на методах нечіткої логіки, наприклад RiskShield.

Байєсівські мережі дають змогу відображати причинно-наслідкові зв'язки між різними впливовими факторами ризику і змінами середовища. На відміну від регресійних моделей БМ дозволяють враховувати не лише безпосередні залежності рівня ризику від факторів ризику, а й залежності між факторами ризику. Крім того, цей клас моделей надає більше можливостей для формування висновку на підставі неповних даних. Із математичного погляду БМ — орієнтований граф, де вершинам відповідають фактори ризику і зміни середовища, а ребрам — виявлені або передбачувані взаємозв'язки [9–13]. Мережа також описується множиною умовних розподілів випадкових величин, що характеризують фактори ризику і змінні середовища. Нехай існує n випадкових змінних X_1, \dots, X_n . Спільну ймовірність розподілу значень цих змінних можна виразити через добуток n умовних ймовірностей:

$$P(x_1, \dots, x_n) = P(x_1) \prod_{j=2}^n P(x_j | (x_1, \dots, x_{j-1})).$$

Перевагою БМ є можливість одночасного використання експертного оцінювання (наприклад, для оцінювання структури мережі шляхом визначення залежностей між змінними, вибір типів апріорних розподілів для змінних) і математичних методів для отримання висновку за мережею. За рахунок цього модель дає змогу зв'язувати вибірки статистичних даних з експертними знаннями.

Висновок на основі БМ можна сформулювати через поширення інформації у будь-якому напрямку. Байєсівські мережі використовують для формування ймовірнісного висновку — розрахунку умовної ймовірності отримання значень для частини випадкових величин за умови відомих значень інших величин. Математично цю задачу можна сформулювати як обчислення $P(y, x)$, де X — множина спостережуваних значень; Y — множина змінних, які потрібно оцінити. Таким чином, задача зводиться до пошуку умовних ймовірностей за формулою [10, 11]

$$P(y|x) = \frac{\sum_{y,s} P(y,x,s)}{\sum_{y,s} P(y,x,s)}$$

де S — множина всіх змінних, крім X і Y . Такі обчислення є трудомісткими, а задача формування висновку NP -повною. Для забезпечення високої точності та спрощення обчислень розроблено й апробовано ряд алгоритмів побудови висновку за БМ.

Залежно від повноти інформації, наявної у суб'єкта підприємницької діяльності, методи оцінювання ризиків умовно можна поділити на групи з урахуванням таких умов:

– *визначеності*, коли інформація про ситуацію ризику достатньо повна, наприклад, у вигляді бухгалтерського балансу, звіту про прибутки, збитків та ін.;

– *часткової невизначеності*, коли інформація про ситуацію ризику існує у вигляді частот появи ризикових подій;

– *повної невизначеності*, коли інформації про ситуацію ризику немає, але є можливість залучення фахівців та експертів для часткового подолання невизначеності.

Таким чином, зменшення обсягу вихідної інформації про середовище діяльності ускладнює оцінювання ступеня ризику через зменшення можливостей та зниження достовірності отриманих результатів.

РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ

Експериментальне дослідження ефективності запропонованої методики виконано за допомогою фактичних статистичних даних. Обсяг статистичної вибірки становить 247 вимірів, які містять такі змінні: назву СК; грошовий еквівалент страхових виплат; статистичний рік; кількість договорів, укладених СК; страхові платежі; кількість страхових випадків за рік. Основна залежна змінна — страхові виплати — яка відображає здійснення грошових переказів з настанням страхового випадку. Решта змінних, включених у вибірку, є незалежними і беруться до уваги як фактори. Структуру статистичних даних подано в табл. 2.

Таблиця 2. Структура статистичних даних

№ з/п	Назва змінної	Характеристики
1	Загальний розмір вибірки	247
2	Статистичний рік	Період 2003–2016 рр.
3	Назва страхової компанії	Текстова змінна
4	Страхові виплати	Поточна сума страхових виплат, грн
5	Кількість страхових договорів	Ціле число, починаючи з 1
6	Страхові платежі	Поточна сума страхових платежів, грн
7	Кількість страхових випадків за рік	Ціле число, починаючи з 1
8	Частота страхових випадків	Задана величина, %
9	Темпи зростання капіталу	Задана величина, %

Як додаткові незалежні змінні використано частоту страхових випадків і темпи зростання капіталу СК. Обидві змінні вимірюються у відсотках і не корелюють із залежною змінною «страхові виплати». Для виконання аналізу

статистичних даних та реалізації алгоритму оброблення екстремальних значень використано такі програмні продукти: інструментальне середовище програмування R2.9.2 для статистичного оброблення даних та роботи з графікою; економетричний пакет Eviews 8.0 для побудови моделей та попереднього оцінювання невідомих параметрів. У пакеті Eviews 8.0 використано такі модулі: розрахунок описових статистик, побудову УЛМ, метод максимальної правдоподібності для оцінювання параметрів моделі. У середовищі програмування R2.9.2 виконано інтеграцію модулів Rcmdr, extRemes, evdbaues та mcmcPack.

У результаті практичного застосування запропонованої вище методики аналізу та прогнозування актуарних процесів побудовано 6 узагальнених лінійних моделей, які відрізняються розподілами залежної змінної (страхові виплати) та видом функції зв'язку. Для подальшого аналізу адекватності побудованих моделей у табл. 3 наведено характеристики кожного з дослідів з використанням середнього значення, стандартного відхилення, інформаційного критерію Акайке, відносної похибки та ін.

Таблиця 3. Характеристики побудованих моделей

Критерій / моделі	Роки	Гамма-логарифмічна	Гамма-тотожна	Пуассона-квадратична	Пуассона-логарифмічна	Обернено-гаусівсько-тотожна	Гауса-логарифмічна
	2003	5956,567	1596,049	1903,869	14916,1	1587,371	24009,31
	2004	11861,295	3904,354	6281,569	16694,01	4425,157	26919,5
	2005	16877,358	22079,792	15667,493	21740,06	17926,272	34320,73
	2006	20923,09	29895,616	21654,081	25026,9	24726,812	37800,2
	2007	22771,274	22403,975	18771,973	24340,78	21963,72	33729,53
	2008	54390,238	74479,367	68189,807	52402,72	61641,887	64950,78
	2009	35443,742	76409,957	60514,376	48730,2	60285,107	65039,51
	2010	38672,291	78164,192	63422,485	51836,74	63034,589	65412,27
	2011	64922,577	93430,554	88418,321	73063,17	79040,921	76376,65
	2012	97963,514	86874,603	90318,79	82214,12	81797,671	66440,83
	2013	118690,588	101936,093	108544,80	114782,69	98590,685	72200,03
	2014	540956,727	200852,822	325641,11	346712,41	158528,148	324616,07
	2015	468048,315	196525,368	311938,27	308718,25	153129,636	315676,9
	2016	23943,557	40046,852	28821,33	28910,16	32357,339	42790,09
	2017	287910,00	163050,00	193736,00	199781,00	51820,00	266978,00
	2018	249921,00	309798,00	173053,00	176268,00	354236,00	137645,00
	Середнє значення	108672,94	73471,40	86434,88	86434,88	61359,67	89305,89
	Стандартне відхилення	171411,80	62670,63	104065,74	106348,93	49938,30	99396,10
Інформаційний критерій Акайке	23,84	325,00	Inf	Inf	324,58	25,30	
Байєсівський інформаційний критерій	BIC-equivalent for q in (0,0028, 0,6772)	BIC-equivalent for q in (0,0293, 0,7131)	CBD	CBD	CBD	BICq equivalent for q in (0,4989, 0,7687)	
Відносна похибка, %	3,14	2,67	2,42	2,65	2,82	3,14	
Величина ризику втрат СК, %	11,27	6,09	8,60	8,79	5,81	7,95	
Скорингова функція Фішера	4	10	5	5	6	13	

Для того щоб визначити, яка з побудованих моделей є найкращим наближенням експериментальних даних до реальних, проаналізовано два класи інформаційних критеріїв: ті, що ґрунтуються на мінімізації відстані Кульбака–Лейблера (ступінь відхилення моделі від реального об’єкта), і ті, що ґрунтуються на байєсівському підході. Результати і порівняння побудованих УЛМ з МГВА показано на рис. 1, 2.

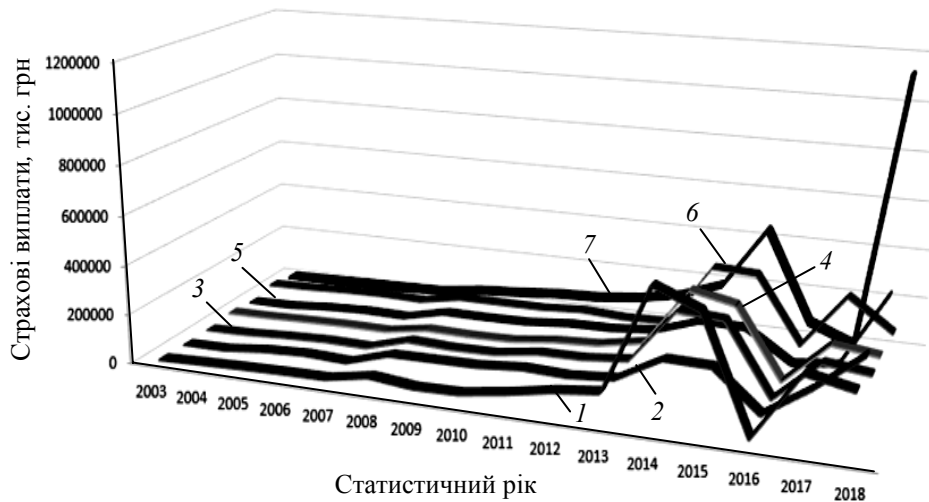


Рис. 1. Порівняння УЛМ з МГВА: 1 — Gamma-log; 2 — Gamma-identity; 3 — Poisson-sqrt; 4 — Poisson-log; 5 — Inverse-gaussian-identity; 6 — Normal-log; 7 — Damages

ID	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	+1	+2
Actual	1586.3	4587.8	8778	12059.5	17210.4	40299	51446.8	66679	72612.8	91535.9	126257	193406	448302	75328		
Predictions				9577	14071	32481	57425	28888	48947	89068	187712	280184	281340	159821	0	1127446
Residuals				-2482.5	-3139.4	-7818	5978.2	-37791	-23665.8	-2467.9	61454.6	86778.1	-166962	84493		

Рис. 2. Результати прогнозування за допомогою МГВА

У результаті порівняння результатів прогнозування, отриманих за допомогою УЛМ і МГВА встановлено, що МГВА дає дещо нижчі результати прогнозування у випадку, коли розподіл даних не відповідає нормальному. Але він дає кращі результати прогнозування за критерієм середньої абсолютної похибки (у відсотках) у випадку, коли дані відповідають нормальному розподілу (табл. 4). Використано НМГВА для прогнозування і за його допомогою отримано найкращий результат, оскільки він не залежить від розподілу.

Після оцінювання втрат, зумовлених страховими виплатами за полісами страхування наступним важливим завданням аналізу актуарних процесів є оцінювання операційного ризику. Для розв’язання цієї задачі побудовано модель у формі мережі Байєса, структуру якої показано на рис. 3.

Функціонування такої мережі апробовано на прикладах з використанням фактичних статистичних даних, а саме: рейтингів СК України за договорами страхування життя у період 2003–2016 рр. Результати оцінювання параметрів БМ подано на рис. 4.

Таблиця 4. Порівняльна таблиця результатів прогнозування нормованого процесу надходження премій

Модель	Сума квадратів похибок прогнозування SSE	Коефіцієнт детермінації R^2	Статистика Дарбіна-Уотсона DW	Середньо-квдратична похибка, %	Середньо-абсолютна похибка, %
AP(2)	4,78099	0,798	2,087	1,445	1,889
AP(6)	4,3953	0,814	1,971	0,928	1,321
АРКС(2;6)	4,39428	0,814	1,999	0,973	1,459
АРКС(6;1)	4,38665	0,815	1,991	0,821	1,235
ЕЗ(0.1)	6,06242	0,745	0,858	2,154	2,789
Тренд(1)	9,019479	0,037	0,487	1,523	1,976
AP(2)+Tr(1)	3,370058	0,640	2,122	1,239	1,627
МГУА	2,731	0,873	2,057	0,147	0,893
НМГУА	2,119	0,884	1,986	0,138	0,699

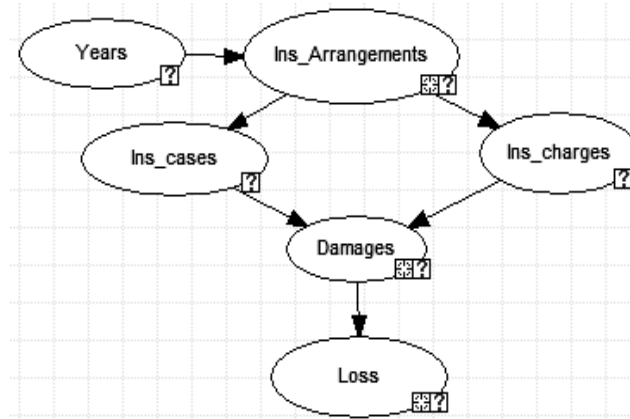


Рис. 3. Графічне зображення БМ довіри для ймовірного оцінювання актуарних процесів

Ranked Targets	Probability
Loss:PresentRisk	0.865
Ins_charges:PaidCharge	0.756
Ins_cases:AbsentCase	0.750
Damages.y2005	0.601
Ins_cases:PresentCase	0.250
Ins_charges:NotPaidCharge	0.244
Loss:AbsentRisk	0.135
Damages.y2006	0.058
Damages.y2004	0.056
Damages.y2009	0.047
Damages.y2003	0.045
Damages.y2011	0.039
Damages.y2010	0.034
Damages.y2007	0.029
Damages.y2008	0.029
Damages.y2014	0.027
Damages.y2012	0.023
Damages.y2013	0.013

Рис. 4. Результати оцінювання БМ

У результаті побудови БМ отримано такі результати: ризик банкрутства СК у випадку 25%-ї ймовірності настання страхового випадку та 100%-ї виплати страхової премії становить 87% за умови, що 78,2% страхувальників вчасно здійснюють платежі за договорами про страхування. Значна величина ризику банкрутства СК свідчить про відсутність ефективного механізму управління коштами як власного капіталу, так і отриманих від страхових договорів. Динаміку страхових виплат у випадку настання страхового випадку за 100%-ї сплати страхових платежів зображено на рис. 5.

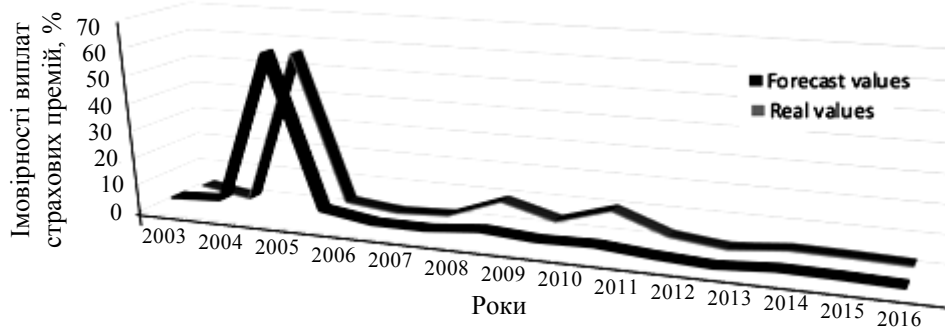


Рис. 5. Динаміка оцінювання страхових виплат

Із графічного зображення динаміки страхових виплат (рис. 5) видно, що прогнознi значення та апріорні ймовірності розрізняються несуттєво завдяки логічній структурі мережі Байєса та високій точності обчислень у мережі. Для навчання структури БМ згенеровано 248 експериментів з випадковими допустимими значеннями щодо кожної з вершин і використано ЕМ-алгоритм (максимізації математичного сподівання) для знаходження (локально) оптимальної максимальної ймовірності оцінок параметрів. Структуру БМ скорочено за допомогою послідовного вилучення вузлів та інверсії ребер початкового графу. Для оцінювання параметрів отриманої структури БМ обчислюється величина операційного ризику СК за 33,9%-ї або 44,2%-ї виплат страхової премії у разі настання (або відсутності) страхових випадків за договорами, за якими надходили/не надходили регулярні платежі. Порівняльні результати наведено в табл. 5.

Таблиця 5. Результати оцінювання спрощеної БМ

CASES	Present insurance case and Paid Charges	Absent insurance case and NotPaid Charges
Pr(Loss)	0,881	0,5
1-Pr(Loss)	0,119	0,5
Pr(Damages)	0,339	0,442

Із табл. 5 видно, що ризик є високим навіть для випадку з 34%-ю виплатою страхових премій СК у разі настання страхових випадків та 100%-ю вчасною сплатою платежів. Рівнозначні значення ризику отримано для протилежного випадку — не було страхового випадку і вчасно не сплачені платежі, тобто договір розірвано. У такому випадку БМ відображає «відсутність ризику» на заданій множині станів, оскільки величина виплати премій досить низька.

У результаті використання запропонованої методики моделювання та створення нових математичних моделей побудовано нові УЛМ та успішно розв'язано задачу оцінювання фінансових ризиків СК. Для вибору й обґрунтування кращої моделі виконано порівняння результатів побудов УЛМ із МГВА та НМГВА.

Для оцінювання ризику банкрутства СК побудовано БМ, функціонування якої апробовано із залученням фактичної статистичної вибірки. Залучення новітніх методів ІАД до розв'язання задачі актуарного моделювання та оцінювання фінансових ризиків відкриває нові можливості для дослідження особливостей методів математичного моделювання фінансових процесів у страхуванні.

ВИСНОВКИ

Виконано дослідження можливості застосування методів ІАД до розв'язання задачі моделювання актуарних процесів та оцінювання ризиків СК. Установлено, що існують методики побудови моделей для опису та оцінювання актуарних процесів, але не достатньо висвітлено питання моделювання актуарних процесів з використанням сучасних методів ІАД. Саме тому запропоновано використання методики моделювання та створення нових математичних моделей на основі використання структур узагальнених лінійних моделей та оцінювання фінансових ризиків СК за допомогою БМ. Для застосування запропонованої методики зібрано необхідні статистичні дані. Під час попереднього аналізу статистичної вибірки виявлено підмножину вироджених даних. Для подальшого опрацювання цих даних розв'язано задачу інтелектуального оброблення екстремальних значень. Оцінювання невідомих параметрів моделі виконано із застосуванням методів максимальної правдоподібності та байєсівського підходу, що належать до ІАД.

У результаті виконання обчислювальних експериментів побудовано нові УЛМ для обраних актуарних процесів, які забезпечують оцінювання високоякісних короткострокових прогнозів щодо страхових виплат. Установлено, що кращою у даному випадку є модель з гамма-розподілом та логарифмічною функцією зв'язку, результат якої отримано за 4 ітерації алгоритму оцінювання параметрів моделі. Проаналізовано якість побудованих моделей у формі УЛМ з використанням множини статистичних критеріїв. Установлено, що такий підхід до моделювання є перспективним, оскільки модель демонструє високі показники якості. Відносні похибки оцінок прогнозів є незначними і коливаються в межах 5 – 12% для всіх моделей, а за допомогою інформаційного критерію Акайке та Байєса обґрунтовано й обрано кращу модель. Отже, застосування УЛМ є ефективним підходом до розв'язання задач прогнозування та оцінювання актуарних процесів.

Розроблено мережу Байєса для ймовірнісного оцінювання операційних ризиків СК. Функціонування такої мережі апробовано на прикладах з використанням фактичних статистичних даних, зокрема рейтингів СК України за договорами страхування життя у період 2003–2016 рр. Отже, ризик банкрутства СК у випадку 25%-ї ймовірності настання страхового випадку та 100%-ї виплати страхової премії становить 87% за умови, що 78,2% страховальників вчасно здійснюють платежі за договорами страхування. Значна

величина ризику банкрутства СК свідчить про відсутність ефективного механізму управління коштами як власного капіталу, так і надходжень, отриманих від страхових договорів. Отже, БМ — це потужний і ефективний математичний інструмент ІАД для дослідження та відтворення реальної картини процесів у СК, який доцільно застосовувати для розв'язання задач імовірнісного прогнозування та оцінювання ризиків. У майбутніх дослідженнях варто вдосконалити структуру БМ та розробити методіку формування висновку в БМ з урахуванням особливостей процесів аналізу та управління операційними ризиками.

ЛІТЕРАТУРА

1. *Jones M.T.* Artificial Intelligence: a systems approach / M.T. Jones. — Hingham, Massachusetts: Infinity Science Press LLC, 2008. — 516 p.
2. *Munakata T.* Fundamentals of the New Artificial Intelligence / T. Munakata. — London: Springer, 2008. — 266 p.
3. *Koski T.* Bayesian Networks / T. Koski, J.M. Noble. — Chichester: John Wiley & Sons, Ltd, 2009. — 368 p.
4. *Darwiche A.* Modeling and Reasoning with Bayesian Networks / A. Darwiche. — Cambridge: Cambridge University Press, 2009. — 562 p.
5. *Дубініна С.В.* Методика аналізу екстремальних даних та її використання при оцінюванні параметрів узагальнених лінійних моделей / С.В. Дубініна, П.І. Бідюк // *Радіоелектроніка, інформатика, управління.* — 2016. — № 1.
6. *McCullagh P.* Generalized Linear Models / P. McCullagh, J.A. Nelder. — New York: Chapman & Hall, 1990. — 526 p.
7. *Tripp M.H.* Quantifying operational risk in general insurance companies / M.H. Tripp // Working Paper [Presented to the Institute of Actuaries] (22 March 2004). — 137 p.
8. *Besag J.* Markov Chain Monte Carlo for Statistical Inference / J. Besag // Working Paper, Center for Statistics and the Social Sciences. — 2001. — N 9. — 25 p.
9. *Дубініна С.В.* Застосування мереж Байєса до побудови моделей оцінювання ризику актуарних процесів / С.В. Дубініна, П.І. Бідюк // *Scientific Journal ScienceRise.* — 2016. — 8, № 2 (25).
10. *Довгий С.О.* Системи підтримки прийняття рішень на основі ймовірнісно-статистичних методів / С.О. Довгий, П.І. Бідюк, О.М. Трофимчук. — К.: Логос, 2014. — 419 с.
11. *Bidyuk P.I.* Forecasting based on Bayesian type models / P.I. Bidyuk, O.M. Trofymchuk, O.P. Gozhyj // *International Journal of Computers and Technology.* — 2015. — Vol. 15, N 3. — P. 6570–6584.
12. *Murphy K.* A Brief Introduction to Graphical Models and Bayesian Networks [Електронний ресурс] / K. Murphy // USA. — 1998. Режим доступу: URL: <http://www.ai.mit.edu/~murphyk/Bayes/bnintro.html>.
13. *Negnevitsky M.* A guide to intelligent systems / M. Negnevitsky. — London: Addison Wesley, 2005. — 435 p.

Надійшла 27.10.2016