Издательство Юрайт. 2024. [Электронный ресурс]. Режим доступа:: https://urait.ru/bcode/540668 (дата обращения: 23.02.2024).

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ, ПРИМЕНЯЕМЫЕ ДЛЯ АНАЛИЗА ФИНАНСОВЫХ ОПЕРАЦИЙ

Ильин Глеб Романович[1]
Российская Федерация, г. Самара, Самарский университет.

Аннотация: Статья посвящена исследованию методов машинного обучения, применяемых для анализа финансовых операций. Описываются основные подходы и алгоритмы, используемые для детектирования подозрительных операций. Приводятся примеры применения этих методов на практике и обсуждаются их преимущества и недостатки. В заключении делаются выводы о наиболее эффективных алгоритмах и предлагаются решения по внедрению данных алгоритмов.

Ключевые слова: машинное обучение, классификация, деревья решений, градиентный бустинг, Anomaly Detection, глубокое обучение.

MACHINE LEARNING APPROACHES TO ANALYSING FINANCIAL TRANSACTIONS

[1]Студент 1 курса магистратуры Института информатики и кибернетики Самарского университета. Научный руководитель: Ростова Е.П., доктор экономических наук, доцент, и.о. заведующего кафедрой математики и бизнес-информатики Самарского университета.

Ilin G.R.
Russian Federation, Samara, Samara University.

Abstract: The article is devoted to the study of machine learning methods used to analyse financial transactions. It describes the main approaches and algorithms used to detect suspicious transactions. Examples of application of these methods in practice are given and their advantages and disadvantages are discussed. The paper concludes with conclusions about the most effective algorithms and suggests solutions for implementation of these algorithms.

Key words: machine learning, classification, decision trees, gradient boosting, Anomaly Detection, deep learning.

Введение

Выявление мошеннических действий стало одной из наиболее актуальных проблем, которая привлекает большое внимание как практиков, так и ученых. В таких отраслях, как банковское дело, ущерб от деятельности киберпреступников в 2023 году по данным ЦБ РФ составил 15,8 млрд рублей [1]. Объемы подозрительных операций с признаками вывода денежных средств за рубеж в первом полугодии 2023 года по данным ЦБ РФ составили 12 млрд рублей [2]. Мошеннические действия могут совершаться при различных операциях: использовании кредитных и дебетовых карт, страховании и т.д. Для решения проблемы выявления мошенничества в рамках переводов по кредитным и дебетовым картам, банки все чаще обращаются к методам машинного обучения, известным своей эффективностью во многих задачах классификации, особенно при работе с неструктурированными данными [3].

Ход исследования

В контексте машинного обучения мошенническая операция представляет собой аномалию. Поскольку модель обучается на данных, где подавляющее большинство транзакций является законными, то мошеннические операции являются аномалиями. Обнаружение этих аномалий в данных о финансовых операциях с использованием инструментов машинного обучения называется Anomaly Detection. Anomaly Detection зачастую представляется в виде классификации.

В математическом виде задача классификации определяется следующим образом: при наличии тренировочной (обучающей) выборки представленной математическим выражением:

$$\left\{\left(x^{(i)}, y^{(i)}\right)\right\}_{i=1}^{n},\ x^{(i)} \in \mathbb{R}^{p}, y^{(i)} \in \{C_1, \dots, C_K\} \qquad (1)$$

следует построить модель $f(x)$, предсказывающую значение каждого объекта. Целевой является функция потерь $L(x, y, f)$, которую необходимо минимизировать, где y – вектор значений целевой переменной, x – матрица объекты-признаки, в которой i–ая строка - это вектор признаков i–го объекта выборки, C – классы объектов.

В процессе Anomaly Detection обнаружение положительных объектов является критически важным. При обнаружении аномалий идентификация операции как положительного объекта (мошенническая операция) может повлечь за собой значительные репутационные и финансовые потери для финансовой организации. В то же время отрицательные объекты очень важны для обучения, и поэтому отметка законной операции как отрицательного объекта – это упущенная возможность для бизнеса.

Две важнейшие особенности выявления подозрительных операций создают проблемы для построения эффективных и точных классификаторов: высокая несбалансированность классов и смещение распределения. В некоторых особо экстремальных случаях можно наблюдать дисбаланс классов, где положительный класс может составлять 1%. Существует ряд подходов, в машинном обучении которые применяются в Anomaly Detection.

Одним из таких является одноклассовая классификация. В отличие от традиционных задач классификации в машинном обучении, подходы одноклассовой классификации (OCC) направлены на выявление образцов, которые не принадлежат к определенному классу. Эти дискриминационные модели “запоминают” границу принятия решения, используя только образцы из определенного нормального класса, что позволяет обойтись без прямой оценки распределения класса. На этапе вывода объекты классифицируются как «объект нормального класса», либо нет, не делая никаких предположений о классе аномалий. Наиболее популярными моделями OCC являются одноклассовый SVM (OCSVM) и Support vector data description (SVDD).

В OCC также используются подходы, основанные на решающих деревьях такие, как isolation forest (IForest) расширенный лес изоляции, Robust Random Cut Forest (RRCF) и PIDForest. Другие методы опираются на зависимости sample-sample (выборка-выборка) для выявления аномалий, например, TracInAD, опирающийся на меры влияния или подходы, основанные на k-nearest neighbors (KNN), в данном методе аномалии выявляются путем измерения расстояния каждого объекта до его k-ближайших соседей, большее расстояние указывает на аномалию.

Методы реконструкции также могут использоваться для обнаружения аномалий. Они основаны на предположении, что выборки и аномалии, генерируемые различными распределениями, распределены нормально. Один из наиболее распространенных методов обнаружения аномалий на основе методов реконструкции, использующих анализ главных компонент (PCA) или байесовский PCA. Автокодировщики, регуляризованные автокодировщики, такие как вариационные автокодировщики (VAE) и глубокие автоэнкодеры с расширенной памятью также используются для обнаружения аномалий.

Также широко применяются *GBDT методы*, основанные на градиентном бустинге на решающих деревьях (Gradient Boosted Decision Trees) или GBDT. В большинстве сценариев используются такие подходы, как XGBoost, LightGBM и CatBoost. GBDT остается наиболее популярным подходом среди практиков благодаря высокой эффективности классификации и простоте обучения по сравнению с методами глубокого обучения [4]. Более того, модели GBDT, такие как LightGBM, XGBoost и CatBoost часто считаются особенно подходящими для несбалансированных и крайне несбалансированных выборок, поскольку эти модели фокусируются на особенно трудно классифицируемых объектах, как правило, классах меньшинства, и таким образом, обеспечивают высокую производительность по сравнению с другими моделями классического машинного обучения. В последние годы именно алгоритмы GBDT показывают себя как достаточно конкурентоспособные алгоритмы на соревнованиях по машинному обучению благодаря своей производительности и скорости [5].

Модель GBDT может быть широко использована даже в компаниях, не обладающих значительными вычислительными

мощностями. В сравнении с нейросетевыми моделями GBDT показывает заметно большую точность, а также, как было замечено выше, модель заметно менее требовательна к вычислительным мощностям. Вследствие этого именно модели GBDT стоит рекомендовать для внедрения.

Полученные результаты и выводы (Заключение)

В ходе сравнительного анализа методов Anomaly Detection было определено, что на данный момент наилучшим методом для задачи выявления мошеннических операций является метод, основанный на градиентном бустинге на решающих деревьях – GBDT. Которые благодаря своей точности и более низким требованиям к вычислительным мощностям (относительно нейронных сетей) являются наиболее оптимальным подходом для выявления мошеннических операций.

Список использованных источников

1) В 2023 году банки предотвратили мошеннические хищения на 5,8 трлн рублей // Банк России. [Электронный ресурс]. Режим доступа: https://www.cbr.ru/press/event/?id=18419 (дата обращения: 28.02.2024).

2) Структура подозрительных операций и отрасли экономики, формировавшие спрос на теневые финансовые услуги // Банк России. [Электронный ресурс]. Режим доступа: https://www.cbr.ru/press/event/?id=18419 (дата обращения: 28.02.2024).

3) Hugo Thimonier, Fabrice Popineau, Arpad Rimmel, Bich-Liˆen Doan 1, Fabrice Daniel Comparative Evaluation of Anomaly Detection Methods for Fraud Detection in Online Credit Card Payments // Universit´e Paris-Saclay, CNRS, CentraleSup´elec. - 2024. - С. 1-17.

4) Gissel Velarde, Anindya Sudhir, Sanjay Deshmane, Anuj Deshmunkh, Khushboo Sharma, Vaibhav Joshi Evaluating XGBoost

for Balanced and Imbalanced Data Application to Fraud Detection // Presented at NVIDIA GTC, The Conference for the Era of AI and the Metaverse. - 2023. - С. 1-20.

5) Marc Wildi, Branka Hadji Misheva A Time Series Approach to Explainability for Neural Nets with Applications to Risk-Management and Fraud Detection // ZHAW Zurich University of Applied Sciences. - 2023. - С. 1-28.

МАКРОЭКОНОМИЧЕСКИЙ И РЕГИОНАЛЬНЫЙ АНАЛИЗ ИНВЕСТИЦИОННОЙ АКТИВНОСТИ В РФ

Кононова Елена Николаевна[1]
Российская Федерация, г. Самара, Самарский университет.
Тикина Анастасия Алексеевна[2]
Российская Федерация, г. Самара, Самарский университет.

Аннотация: В статье рассматривается система показателей инвестиционной активности, применяемых современной российской статистикой и дополнительно предлагаемых авторами для оценки уровня инвестиционной активности на макроэкономическом и региональном уровнях. Проведен анализ состояния обозначенных индикаторов за пятилетний период в РФ и ее регионах, выявлены сложившие тенденции их изменения и проблемные состояния. Они могут быть учтены при

[1]Кандидат экономических наук, доцент, доцент кафедры экономики инноваций Самарского университета.
[2]Студент 2 курса магистратуры Института экономики и управления Самарского университета.