

Прогнозирование дефолта компании с использованием графов юридических лиц

Аннотация

Цель проекта – повысить точность прогнозирования корпоративных дефолтов путем интеграции графового анализа связей юридических лиц в модель кредитного риска. В рамках работы будут построены графы компаний на основе структур собственности, отношений «материнская–дочерняя» компании, транзакционных связей и исторических случаев дефолта. Из графов планируется извлечь информативные признаки (метрики центральности, embedding-и, структуры сообществ) и объединить их с финансовыми показателями компаний. Для прогноза вероятности дефолта будут применены модели машинного обучения (например, градиентный бустинг) и глубокого обучения на графах (Graph Neural Networks, GNN). Оценка качества моделей будет проводиться по метрикам AUC, F1 и PR-кривым, что позволит сравнить эффективность традиционных и графовых подходов. Ожидается, что включение сетевых признаков выявит скрытые взаимосвязи между компаниями и улучшит точность предсказания дефолта по сравнению с моделями, опирающимися только на финансовую информацию.

Введение

Прогнозирование вероятности дефолта компании является ключевой задачей финансового риска, имеющей большое значение для банков, инвесторов и регуляторов. Традиционно для оценки кредитоспособности используются только финансовые показатели и бухгалтерские коэффициенты компании, однако такой подход не учитывает её связь с другими фирмами и контрагентами. При этом компании образуют сложные сети через отношения собственности, совместных руководителей, клиентско-поставщицкие связи и прочие деловые взаимодействия. Исследования показывают, что финансовые проблемы могут распространяться по таким сетям: например, убыточность или неплатежеспособность одного предприятия повышает риск дефолта у связанных с ним контрагентов. В связи с этим возникает гипотеза, что использование параметров, описывающих положение компании в межфирменном графе, позволит более точно оценивать её вероятность дефолта по сравнению с моделями, опирающимися исключительно на собственные финансовые показатели фирмы. Настоящий проект нацелен на проверку этой гипотезы путем интеграции методов анализа сетей юридических лиц в модель прогноза дефолтов.

Обзор литературы

Для повышения качества прогнозирования дефолтов исследователи стали привлекать данные о взаимоотношениях компаний и применять методы сетевого анализа. Так, Tobback с соавт. построили сеть малых и средних предприятий по общим директорам и с помощью специального сетевого классификатора рассчитали для каждой фирмы «реляционный» скор на основе дефолтов её связанных соседей, добавив этот показатель в модель наряду с финансовыми коэффициентами. В результате комбинированная модель показала более высокую точность по сравнению с использованием только финансовых данных: совместное применение сетевых и финансовых переменных дало наибольший прирост качества прогноза. Другой подход предложен Martínez и др., которые анализировали граф связей «поставщик–покупатель»: расчёт метрик графа (например, степени центральности узлов и других структурных характеристик) и применение графовой свёрточной нейросети (GCN) позволили существенно улучшить предсказание будущих дефолтов предприятий.

В последние годы набирают популярность графовые нейронные сети (GNN) для задач кредитного риска, поскольку они учитывают нелинейные зависимости между связанными компаниями, недоступные традиционным моделям. Например, Brogaard и Chen (2024) сообщают, что модель GNN с информацией о межфирменной сети улучшила точность прогноза кредитных спредов более чем на 50% по сравнению с классическими алгоритмами. Minakawa и соавт. (2024) продемонстрировали, что специальный метод отбора соседних узлов в большом графе (с приоритетом связей с большим объёмом транзакций) позволяет повысить стабильность и качество прогнозирования дефолтов по сравнению со стандартным случайным отбором соседей. На практике графовый подход уже применяется в банковской сфере: например, Газпромбанк использует граф юридических лиц (связи через общих руководителей, доли владения и транзакции) для расчёта метрик расстояний и центральностей, которые интегрируются в кредитный скоринг и дают заметный прирост точности по сравнению с моделями без учёта графов.

Описание данных

Для исследования потребуется совокупность финансовых и сетевых данных о компаниях. Финансовые показатели будут включать бухгалтерскую отчётность и коэффициенты (ликвидность, задолженность, рентабельность и др.), характеризующие текущее состояние и динамику бизнеса каждой компании. Сетевые данные представляют собой граф юридических лиц, где узлами являются компании, а рёбра отражают различные виды связей между ними. В частности, будут учитываться отношения владения (доли участия, связи «материнская–дочерняя» компания), пересечения в руководстве или составе акционеров, а также коммерческие взаимоотношения — например, кредитно-долговые или транзакционные связи между фирмами. Кроме того, для некоторых компаний будет отмечено событие дефолта (банкротства) в прошлом, что позволит использовать эти сведения как метки для обучения и как информацию в графе (например, наличие в окружении фирмы узлов с историей дефолта). Данные планируется собрать из открытых источников (базы корпоративной отчетности, реестры юридических лиц) и внутренних баз данных, охватывая достаточное число предприятий для статистически значимых выводов.

Запланированные методы SNA

Для построенного графа компаний будут рассчитаны различные показатели сетевого анализа, в том числе:

- **Метрики центральности узлов** – количественные меры влияния и значимости компании в сети (например, степень узла, посредничество, близость, центральность собственного вектора и др.).
- **Векторные представления узлов (node embeddings)** – обучение представлений вершин графа в виде числовых векторов, сохраняющих структуру связей. Планируется использовать методы вложений графов (например, алгоритм Node2Vec) либо обучать представления с помощью графовых нейронных сетей, чтобы кодировать структурные особенности окружения каждой компании.
- **Обнаружение сообществ** – выявление кластеров тесно связанных компаний в графе (например, алгоритмом Лувена). Принадлежность компании к тому или иному сообществу, размер этого сообщества и другие характеристики могут быть использованы как признаки, отражающие структуру сети.

Извлечённые сетевые признаки затем будут объединены с традиционными финансовыми показателями и использованы в моделях прогнозирования дефолта. В частности, планируется построение ансамблевых моделей машинного обучения (градиентный бустинг деревьев решений) с расширенным признаковым пространством, а также экспериментальное применение графовых нейронных сетей (GNN), способных непосредственно обучаться на структуре графа. При наличии исторических данных за несколько периодов возможна реализация временных моделей, учитывающих эволюцию показателей и сетевых связей во времени.

Ожидаемые результаты

Предполагается, что модели, учитывающие сетевые признаки, продемонстрируют более высокую точность прогноза дефолтов по сравнению с базовыми моделями, основанными только на финансовых данных. Ожидаемый прирост качества будет отражён в повышении показателей AUC, полноты, точности и улучшении PR-кривой для позитивного класса дефолта. Помимо общего улучшения метрик, анализ значимости признаков (например, в модели градиентного бустинга) и интерпретация параметров графовой модели (GNN) позволят выявить, какие сетевые характеристики наиболее сильно коррелируют с риском дефолта. Предполагается, что ключевыми факторами могут оказаться высокое значение центральности компании в корпоративной сети, участие в плотном сообществе связанных фирм или, напротив, изолированное положение, а также наличие прямых или косвенных связей с уже обанкротившимися предприятиями. Выявление таких структур и паттернов в графе даст более глубокое понимание системных причин корпоративных дефолтов и покажет, как использование сетевой информации позволяет точнее идентифицировать организации с повышенным кредитным риском.

Список литературы

1. Савинов А. Графовый анализ — обзор и области применения // Хабр, 2021. — URL: <https://habr.com/ru/companies/glowbyte/articles/594221/> (дата обращения: 20.05.2025).
2. Tobback E., Bellotti T., Moeyersoms J., Stankova M., Martens D. Bankruptcy prediction for SMEs using relational data // Decision Support Systems. — 2017. — Vol. 102. — P. 69–81. — URL: <https://doi.org/10.1016/j.dss.2017.07.004> (дата обращения: 20.05.2025).
3. Martínez A., Nin J., Tomás E., Rubio A. Graph Convolutional Networks on Customer/Supplier Graph Data to Improve Default Prediction // Complex Networks X: Proceedings of the 10th International Conference on

- Complex Networks (CompleNet 2019). — Springer, 2019. — P. 135–146. — URL: https://doi.org/10.1007/978-3-030-14459-3_11 (дата обращения: 20.05.2025).
4. Zhao J., Ouenniche J., De Smedt J. A complex network analysis approach to bankruptcy prediction using company relational information-based drivers // Knowledge-Based Systems. — 2024. — Vol. 300. — Article 112234. — P. 1–24. — URL: <https://doi.org/10.1016/j.knosys.2024.112234> (дата обращения: 20.05.2025).
 5. Minakawa N., Izumi K., Murayama Y., Sakaji H. Firm Default Prediction by GNN with Gravity-Model Informed Neighbor Node Sampling // The Review of Socionetwork Strategies. — 2024. — Vol. 18. — P. 303–328. — URL: <https://doi.org/10.1007/s12626-024-00170-6> (дата обращения: 20.05.2025).
 6. Brogaard J., Chen B. Attention-based Graph Neural Networks in Firm CDS Prediction. — 2024. — (Preprint). — URL: https://www.belindachenchen.com/uploads/GNN/Attention_based_Graph_Neural_Networks_in_Firm_CDS_Prediction.pdf (дата обращения: 20.05.2025).