

## Доклад на семинаре «Банки и предприятия: модели и рейтинги»

**Название доклада:** «Предсказание кредитного рейтинга на основе самоорганизующихся карт».

**Докладчик:** Пастухов Евгений Стахович, канд. физ.-мат. наук, Главный экономист АКБ «ЦентроКредит».

**Цель доклада:** привлечение внимание банковских аналитиков к одному из существующих подходов к построению систем рейтингования компаний с использованием нейронно-сетевых средств моделирования.

**Краткое содержание доклада.** Доклад посвящен изложению и оценке научно-прикладного исследования «Предсказание кредитного рейтинга на основе самоорганизующихся карт», выполненного в 2000г. в Голландии на экономическом факультете Роттердамского ун-та, Роджером Таном (Roger Tan). Исследование можно отнести области финансовой эконометрики и нейровычислений.

*Мотивировка исследования.* Установление рейтинга компании в отношении ее кредитоспособности является важнейшим событием, как для самой компании, так и для ее возможных инвесторов. Кредитные рейтинги используются в качестве меры риска инвестиций в долговые обязательства и как определенные пределы объема покупки облигации некоторым инвестором. Так для инвестиционных банков существует рейтинговое ограничение на приобретение активов только с инвестиционным рейтингом (первые 4 градации).

Однако процесс рейтингования авторитетными агентствами, такими как Standard & Poor's и Moody's, является дорогостоящей и относительно длительной процедурой. Более того, поступление существенно новой информации о компании требует, вообще говоря, постоянного пересмотра рейтинга. В такой ситуации у банков, заинтересованных более других инвесторов в качественной и своевременной оценке кредитоспособности эмитента ценных бумаг, возникает желание в создании собственных систем, использующих те же критерии, что и рейтинговые агентства. Если результаты оценки кредитоспособности обеими системами сопоставимы, то кредитор экономит время и деньги, не прибегая к дорогостоящим услугам рейтинговых агентств.

*Постановка задачи.* Создание системы-имитатора кредитного рейтинга является трудным занятием, так как агентства тщательно охраняют алгоритм присвоения рейтинга, ссылаясь лишь на используемые финансовые показатели и качественные факторы. Поскольку качественные оценки заемщика проводятся экспертами агентств и не известны публично, то единственным источником информации для построения рейтинговой системы являются общедоступные данные финансовой отчетности.

Рейтинговые агентства неоднократно подчеркивали, что львиную долю в комплексной оценке рейтинга эмитента занимают субъективные оценки экспертов, проводящих наблюдения и опросы сотрудников, например, их мнения о качестве долгосрочной стратегии, существовании будущих угроз потери прибыли и т.п., что естественно никак не отражается в финансовой отчетности компании.

В этой связи возникает ключевой вопрос, в какой степени рейтинг зависит от финансовых показателей, чтобы только на них основывать построение системы-имитатора?

*Существующие подходы.* Для построения системы рейтингования, учитывающей только финансовые данные эмитента, предпринимались многочисленные попытки. В их основе лежали регрессионные методы, среди которых наибольшей точностью отличались прямоточные нейронные сети, представляющие собой класс нелинейных непараметрических регрессий. Анализ результатов на эту тему показал, что нейросетевые оценки рейтинга точнее традиционных линейно-регрессионных примерно на 10-20%. Вместе с тем, принципиальным недостатком упомянутых прямоточных нейронных сетей

является их полная «закрытость» (модель «черного ящика»), что делает практически невозможным содержательную экономическую интерпретацию полученных с ее помощью результатов.

В отличие от прямоточных нейронных сетей, самоорганизующиеся карты признаков Кохонена (Self-Organizing Maps, далее СОК) представляют собой особый класс нейроно-сетевых парадигм, способных отображать тренировочное и тестовое множества примеров в своем выходном слое нейронов, как на своеобразном дисплее или карте, сохраняя при этом топологическую и кластерную структуры исходного пространства обучающих данных. В результате у исследователя появляется возможность полной интерпретации исходных данных и полученных с их помощью результатов за счет двухмерной визуализации многомерных данных.

*Результаты исследований.* В упомянутом исследовании сделаны попытки обнаружить статистически значимую связь между данными финансовой отчетности и кредитным рейтингом компании, присвоенным агентством Standard & Poor's, а также построению на реальных данных множества компаний двух моделей:

- дескриптивной (описательной) модели кредитного рейтинга, основанной только на финансовых показателях;
- классификационной модели, основанной на финансовых данных и соответствующих им рейтинговых оценках.

Исследование состоит из 4-х частей и Заключения, содержащего итоговые выводы.

*Первая часть* посвящена описанию процесса присвоения кредитного рейтинга агентством Standard & Poor's, в том числе, интерпретации рейтинговой шкалы и финансовых коэффициентов, учитываемых в процедуре рейтингования. Кроме этого, в качестве подтверждения значимости рейтинга, приводится принципиально важная диаграмма соответствия шкалы кредитного рейтинга и реально существовавших дефолтов оцениваемых компаний.

*Вторая часть* полностью посвящена описанию возможностей метода СОК как адекватного аналитического инструментария. Во-первых, СОК рассматривается как разновидность (наряду с методом главных компонент и многомерного масштабирования) проекционных методов, позволяющая сохранять исходный порядок и визуализировать с помощью системы карт многомерные данные. Это обстоятельство обосновывает возможность изучения многомерных данных с помощью их двухмерных проекций на картах.

Во-вторых, СОК помогает кластеризовать изучаемые многомерные объекты в однородные группы (классы). Объекты (в нашем случае рейтингуемые компании) внутри данного класса имеют весьма схожие значения финансовых показателей и по логике принципа «схожие компании не могут иметь разные рейтинги» должны быть одинаково оценены.

В-третьих, СОК является полупараметрической нелинейной регрессией, которую можно использовать в качестве классификационной модели для прогнозирования рейтинга новой компании (еще не оцененной) или требуемой оценки на новую отчетную дату.

В четвертых, с помощью СОК можно осуществить нелинейный корреляционный анализ учитываемых для рейтингования показателей, с целью устранения из их списка зависимых и малоинформативных.

В пятых, метод СОК успешно работает с искаженными данными, выбросами и пропусками, а также не требует (для своего корректного применения) проверки теоретико-вероятностных предположений относительно природы источника данных.

*В третьей главе* с помощью СОК строится дескриптивная модель кредитного рейтинга, основанная только на используемых агентством Standard & Poor's 18 финансовых коэффициентах, разбитых на 6 тематических групп: процентного покрытия, леввериджа, рентабельности, размера компании, стабильности и рыночности. В качестве объектов моделирования были выбраны отрейтингованные Standard & Poor's 294 компании потребительского сектора США, данные о которых заимствованы из источника Compustat

Quarterly. Данные относились к 5-ти летней истории наблюдений (1994 –1998гг.) и использовались с точностью до квартала, так как их рейтинг ни разу не менялся дважды в течение двух кварталов (политика рейтинговых агентств держать рейтинг постоянным как можно дольше).

В результате преобработки и трансформации исходных данных была построена дескриптивная модель, которая в процессе многочисленных оптимизаций по финансовым коэффициентам и внутренним параметрам алгоритма привела к формированию 8 кластеров на карте и соответственно 8 классов компаний, каждая из которых получила обоснованную и содержательную (в терминах отобранных финансовых коэффициентов) экономическую интерпретацию. Полученные формальным (оптимизация по коэффициенту качества кластеризации) путем классы компаний оказались согласованы с рейтингом Standard & Poor's (с лагом в два квартала), который не использовался при построении кластерной модели. Под согласованностью понимается корреляция рейтинга компаний со средним рейтингом по классу. В результате сделан вывод о том, что с помощью построенной модели примерно 60% рейтинга компаний может быть объяснено их финансовыми показателями, остальные, по-видимому, 40% приходятся на качественные экспертные оценки. Проведенный анализ чувствительности построенной модели к вариациям обучающих данных и параметров самой модели показал ее высокую стабильность.

В четвертой главе СОК используется как нелинейная регрессия для построения классификационной модели рейтинга. Для этого, как это всегда делается при обучении нейронных сетей, исходный массив данных делится на три случайных множества соответственно для тренировки, оптимизации и окончательного тестирования. В отличие от дескриптивной модели нейронной сети при обучении показывается значения рейтинга в виде числового кодом, сохраняющего порядок исходной шкалы, состоящей из 22 градаций. В качестве критериев качества построенной модели выступали коэффициенты правильной классификации Success ratio, средняя абсолютная ошибка MAD, коэффициент детерминации R<sup>2</sup> и статистическая значимость построенной регрессии. Большая часть исследований заняла оптимизация модели по комбинациям финансовых коэффициентов и многочисленным внутренним параметрам, в результате чего была выбрана оптимальная по точности классификации модель, использующая двухлетний период наблюдений (историческая глубина), всего 5 (из 18) коэффициентов, с размером карты в 1000 нейронов и радиусом соседства в 50 нейронов.

Оценки точности классификации по тестовой выборке выглядят следующим образом: Success ratio = 60% при ошибке в +-1 позиция рейтинга и 82% при ошибке в +-2; MAD = 1,40, R<sup>2</sup> = 0,64. Сравнение построенной модели с моделями линейной регрессии и Ordered logit показало примерно схожие результаты. Интересно отметить, что наиболее высокая точность классификации у всех трех типов моделей наблюдались внутри рейтинговых классов с номерами 9 (B+) по 17 (A). Проверка точности СОК-модели на других исторических этапах подтвердила ее относительную стабильность во времени.

В выводах к проведенному исследованию говорится о принципиальной возможности построения модели кредитного рейтинга компаний, основанной только на использовании данных финансовой отчетности. При этом приблизительно 80% рейтингуемых компаний могут быть оценены с точностью +-2 позиции рейтинговой шкалы агентства Standard & Poor's. Другим важным выводом является тот факт, что данное рейтинговое агентство ошибается, переоценивая значение и вклад качественных факторов, в противоположность количественным (финансовым коэффициентам). Еще одним полезным выводом является то, что построенная модель может служить хорошей аппроксимацией, когда реальный кредитный рейтинг в нужный момент отсутствует.