“Моделирование кредитных рисков коммерческого банка с использованием методов машинного обучения”

Николай Безносов

Виды банковских рисков

- Кредитные риски
- Процентные риски
- Валютные риски
- Рыночные риски
- Риски ликвидности

Кредитный риск - риск возникновения дефолта дебитора

Кредитный риск

	Розничное направление	Корпоративное направление
1) Объект кредитования	Физические лица	Юридические лица
2) Методы оценки рисков	• Экспертные оценки • Статические модели • Машинное обучение	• Оценки рейтинговых агентств • Структурные (рыночные) модели • “Reduced-form” модели • Статистические модели
3) Доступные данные	1) Личная информация о заемщике (возраст, пол, семейное положение и т.п.) 2) Информация о всех кредитах 3) Информация о доходе 4) Информация из социальных сетей	1) Данные отчетности 2) Оценки рейтинговых агентств и аналитиков 3) Биржевая информация (рыночные цены и т.п.) 4) Текстовые данные (новости, отчеты аналитиков и т.п.)

Подходы к оценке вероятности дефолта компании

Группа моделей	Авторы ключевых работ	Краткое описание
Структурные (рыночные)	Black & Scholes (1973) и Merton (1974); Fisher, Heinkel, Zechner (1989); Leland (1994); Vasicek (1977) и Shimko (1993); Schwartz (1995) и Hui et al. (2003)	Вероятность дефолта определяется разницей между рыночной оценкой активов компании и ее обязательствами. Ключевую роль играет волатильность активов, которая оценивается по формуле Блэка-Шоулза.
“Reduced-form”	Jarrow and Turnbull (1992, 1995); Duffie and Singleton (1999)	Дефолт – случайная экзогенная величина, подчиняющаяся Пуассоновскому процессу
Статистические	Beaver (1966, 1968) и Altman (1968); Ohlson (1980)	Показатели бухгалтерской отчетности используются для формирования рейтинга платежеспособности. Чем ниже рейтинг, тем выше вероятность дефолта.

Модель Мертона

Предпосылки:

- Капитал компании – собственные средства (E) и долг (D)
- Нет дивидендов
- Долг в форме облигаций с нулевым купоном с датой погашения T
- Стоимость компании $V_t = D_t + E_t$, где t - время
- В момент времени T все, что осталось после погашения долга уходит акционерам и фирма закрывается
- Если $V(T) < D(T)$, то считаем это дефолтом

Модель Мертона

Заметим, что стоимость E эквивалентна стоимости европейского “call” опциона на V с ценой исполнения D и сроком исполнения T

Условие	Получат кредиторы	Получат акционеры
V(T) > D	D	V(T) - D
V(T) < D	V(T)	0

$C = V * N(d_1) - D * e^{-r*T} * N(d_2)$

$d_1 = \frac{\ln(V/D) + (r+\sigma^2/2)*T}{\sigma\sqrt{T}}$

$d_2 = d_1 - \sigma\sqrt{T}$

r – ставка процента

N(.) – кумулятивная функция распределения стандартного нормального распределения

C – цена “call” опциона

$P(D) = 1 - N(d_2)$

Модель Jarrow и Turnbull

Дефолт – экзогенная переменная, подчиняющаяся процессу с интенсивностью λ(t)

Тогда вероятность дефолта:

$$P[N(T) = 1 \mid N(t) = 0] = 1 - e^{-\int_t^T \lambda(s)ds}$$

Как и в модели Мертона, рассмотрим облигацию с нулевым купоном.

Ее цену представим как матожидание будущей прибыли. В случае дефолта получаем δ, иначе 1.

$$BondP = e^{-r(T-t)} * E[1 * e^{-\int_t^T \lambda(s)ds} + \delta * (1 - e^{-\int_t^T \lambda(s)ds})]$$

Отсюда можем найти λ и вероятность дефолта в период [t, T].

Модель Альтмана

- Данные о 66 предприятиях, половина из которых обанкротились
- Данные о показателях бухгалтерской отчетности
- Использовал дискриминантный анализ для нахождения разделяющей плоскости

$Z = 0.012*X_1 + 0.014*X_2 + 0.033*X_3 + 0.006*X_4 + 0.999*X_5$

X_1 - оборотный капитал / активы

X_2 - прибыль / активы

X_3 - EBITDA / активы

X_4 - Рыночная стоимость капитала / Стоимость долга

X_5 - Выручка / активы

Модель Олсона

- Данные о 2163 компаниях, 105 из которых обанкротились с 1970 по 1976
- Данные о показателях бухгалтерской отчетности
- Логистическая регрессия

Признаки:

1) log(активы / ВНП)
2) Обязательства / активы
3) Оборотный капитал / активы
4) Текущие обязательства / текущие активы
5) Чистая прибыль / активы
6) Оборотные активы / обязательства

The Failure of Supervisory Stress Testing

W. Scott Frame u Kristopher Gerardi (2014)

Тестируют старые модели кредитного скоринга на современных данных

Они выделяют три ключевые причины модельного риска:

1) Скоринговая модель была примитивной и учитывала крайне ограниченное количество факторов

2) Спецификация модели не пересматривалась в течение десятилетий:
   - закрывая глаза на изменения в экономических условиях
   - не учитывая эволюции поведения игроков и рыночных практик
   - игнорируя появление свежих, более детальных и актуальных, данных
   - пренебрегая новыми, более продвинутыми статистическими методами

3) Модель была слишком прозрачной и доступной для всех, в т. ч. для оцениваемых финансовых организаций, которые подстроили свои действия под конкретные метрики

Что делать?

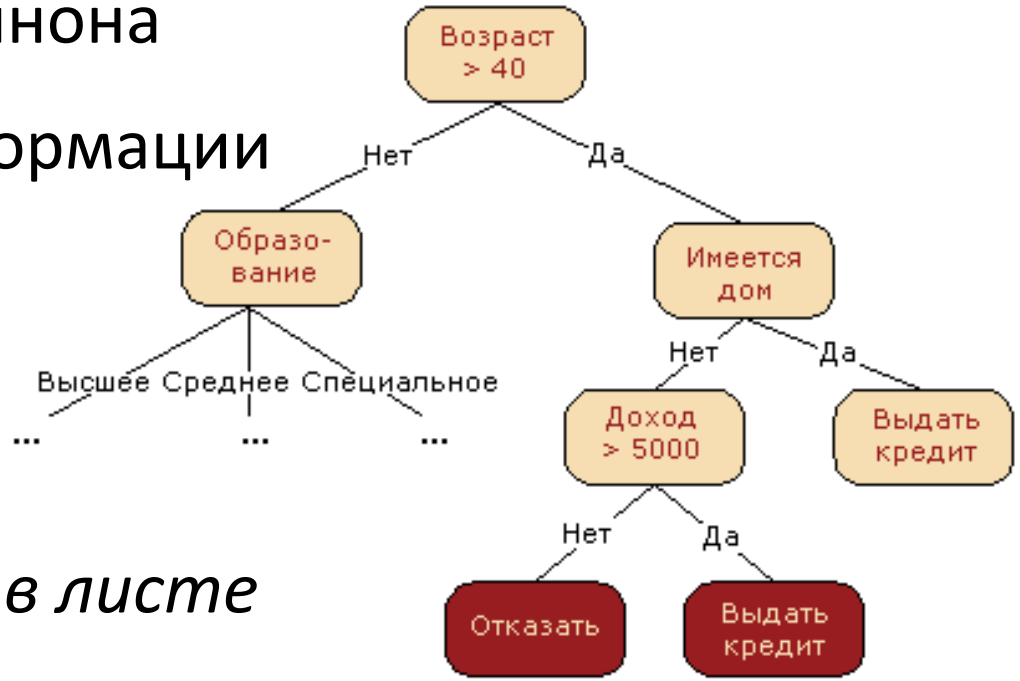
1) Использовать больше данных
2) Использовать данные разных типов в одной модели (отчетность, биржевая информация, оценки аналитиков и рейтинговых агентств, текстовая информация)
3) Использовать более сложные и гибкие модели и их ансамбли
4) Позволить моделям самостоятельно дообучаться на новых данных

Модели машинного обучения

Линейные	Метрические	Древесные	Нейронные сети
Линейная регрессия Логистическая регрессия SVM Наивный Байес	KNN	Decision tree Random forest Gradient tree boosting	MLP RNN LSTM

Дерево решений

- $S = -\sum_{i=1}^{N} p_i log_2(p_i)$ - энтропия Шеннона
- $IG(Q) = S_0 - \sum_{i=1}^{q} \frac{N_i}{N} S_i$ - прирост информации



Контролируем:

- *Глубину дерева*
- *Минимальное количество примеров в листе*

Случайный лес

Впервые введен Tin Kam Ho в 1995 году, улучшен до текущей версии Breiman L в 2001 году.

Алгоритм построения случайного леса, состоящего из N деревьев:
Для каждого n=1,…,N:

1. Сгенерировать выборку X_n с помощью бутстрэпа;
2. Построить решающее дерево b_n по выборке X_n:
   - по заданному критерию мы выбираем лучший признак, делаем разбиение в дереве по нему и так до исчерпания выборки
   - дерево строится, пока в каждом листе не более n_{min} объектов или пока не достигнем определенной высоты дерева
   - при каждом разбиении сначала выбирается m случайных признаков из n исходных,
   и оптимальное разделение выборки ищется только среди них.

Итоговый классификатор $a(x) = \frac{1}{N}\sum_{i=1}^{N} b_i(x)$ задачи классификации мы выбираем решение голосованием по большинству

Контролируем:

- *Глубину дерева*
- *Количество деревьев*
- *Минимальное количество примеров в листе*
- *Число признаков, по которым ищется разбиение*

Градиентный бустинг над деревьями

Идея оптимизации бустингом была предложена Breiman L. в 1997 году. Первая работа по использованию градиентного бустинга для построения ансамбля деревьев была написана Friedman J. в 1999 году.

$$b_1(x) = argmin\ \frac{1}{m}\sum((b(x_i) - y_i)^2$$

$$b_1(x_i) + b_2(x_i) = y_i$$

$$b_2(x) = argmin\ \frac{1}{m}\sum(b(x_i) - (y_i - b_1(x_i)))^2$$

- *Строим базовый алгоритм* b_0
- *Пусть у нас уже есть* a_{N-1} *алгоритмов. Нужно выбрать* $b(x)$ *такой, чтобы:*

$$F = \sum L(y_i, a_{N-1}(x_i) + b(x_i)) \rightarrow min$$

То есть нам нужно найти вектор $s_i = b(x_i)$*, который сильнее всего уменьшает значение* $F(s)$*.*

А это антиградиент функции $F(s)$*.*

- *Строим новый алгоритм по новым y, а именно по антиградиенту F*

$$b_N = argmin\ \frac{1}{m}\sum((b(x_i) - s_i)^2$$

- $a(x) = \sum b(x)$

Предполагаемый вклад

Цель – построение модели корпоративного скоринга.

- Разработка оптимальной архитектуры моделей, способной работать с разнородными данными
- Разработка оптимальной loss-функции применительно к корпоративному кредитному скорингу
- Проверка гипотез о влиянии параметров на вероятность дефолта
- Разработка системы принятия управленческих решений на основе результатов модели
- Выбор оптимального решения ряда проблем задач машинного обучения (см. след. слайд) применительно к корпоративному кредитному скорингу

Возможные проблемы

1) Несбалансированность классов

Решение:

- Модификация функции потерь таким образом, чтобы сильнее штрафовать за неправильную классификацию объектов меньшего класса
- Использование метрик качества, индифферентных к несбалансированности классов

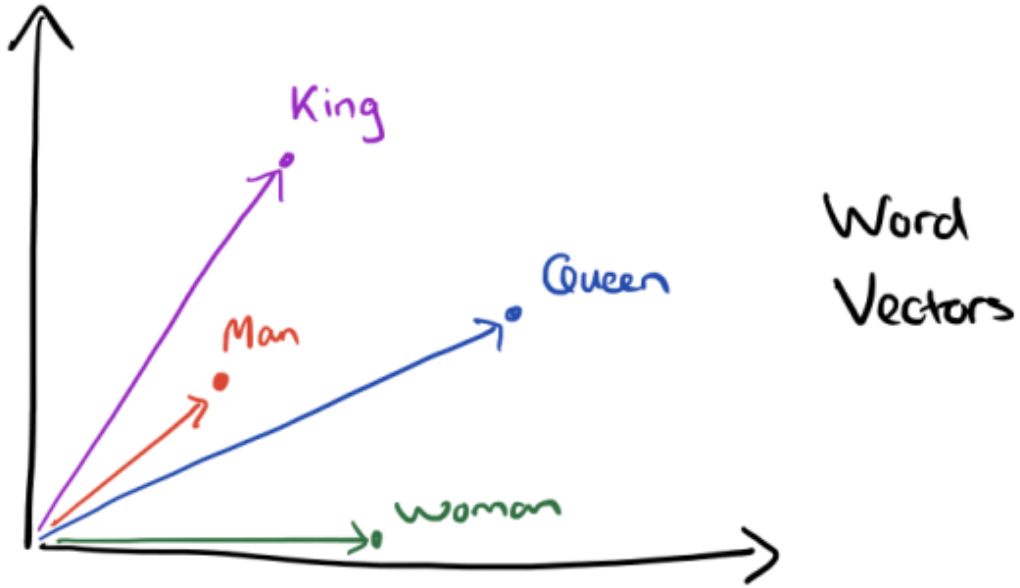
2) Как учесть данные разных типов в одной модели – временные ряды (биржевые цены), точечные данные (отчетность), текстовая информация

- Генерация признаков или нейронные сети
- Для текста использовать word2vec

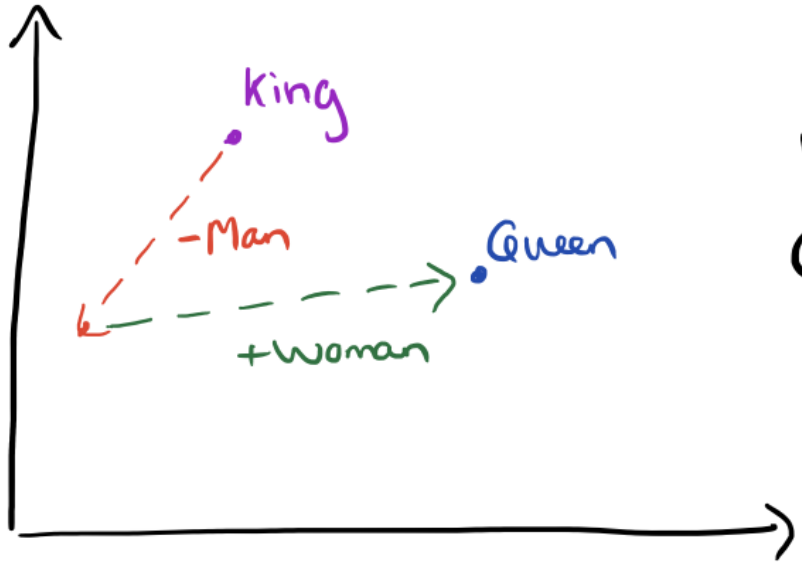
Word2Vec

Технология, разработанная Google в 2014 году.

Метод, позволяющий представлять слова в виде векторов.



Word Vectors



Vector Composition

Спасибо за внимание!