

РАЗМЕТКА ДАННЫХ НА ОСНОВЕ STOP LOSS И TAKE PROFIT ДЛЯ СОЗДАНИЯ ТОРГОВОЙ СИСТЕМЫ

Кузнецов Д. И., студент кафедры информатики СПбГУ,
denis417k@gmail.com

Григорьев Д.А. к.ф.-м.н, доцент кафедры информатики СПбГУ,
gridmer@yandex.com

Аннотация

В данной работе описывается создание торговой системы на основе градиентного бустинга. Задача классификации является одной из самых распространенных в сфере машинного обучения, вследствие чего разработано большое множество алгоритмов обучения с учителем для её решения. Для обучения необходимы размеченные данные, метки которых диктуют, как будут интерпретироваться результаты. В нашей работе разметка основана на предсказании направления исполнения Стоп-заявки. Кроме того, для предсказания используются данные технического анализа.

Введение

Для эксперимента необходимы исторические данные определенного финансового инструмента. Данные представляют собой временной ряд вида OHLC (Open, High, Low, Close).

Технический анализ финансового рынка является одним из основных инструментов прогнозирования изменения цен финансовых инструментов. Он производится над временным рядом OHLC. В результате проведенных экспериментов были отобраны конкретные индикаторы и подсчитаны их соотношения в качестве признаков. Полученные признаки могут быть поданы классификатору для дальнейшего предсказания метки класса.

При создании модели нами была использована библиотека CatBoost от Яндекса, в частности, класс CatBoostClassifier, предоставляющий для настройки множество гиперпараметров. Основные из них - это параметр l2 регуляризации, глубина деревьев, функция потерь.

Описание системы

Признаки

Для обучения модели было отобрано 39 признаков. Стоит заметить, что мы старались исключить из признаков те индикаторы, которые характеризуются исключительно абсолютным значением и, поэтому, их изменение не несет никакой информации. Это необходимо для улучшения качества классификации. Некоторая часть из признаков представлена в таблице 1.

Признак	Описание
ATR	Средний истинный диапазон — показатель волатильности рынка.
EMA(9)	Скользящая средняя с периодом 9.
ADX(14)	Средний индекс направленного движения. Отображает наличие или отсутствие тренда.
MACD_d=MACD-MACD_signal	Пересечение линий MACD и сигнальной линии MACD дает сигнал на покупку или продажу в зависимости от относительного положения линий до пересечения.
Bollinger Bands (BB) - Close	Показывает нахождение цены инструмента относительно полос Боллинджера.
RSI	Индекс относительной силы. Определяет силу тренда и вероятность его смены.
RSI - Close	Схождение и расхождение цены инструмента и RSI дают информацию об окончании тенденции.
Stochastic Oscillator	Показывает положение текущей цены относительно диапазона её цен за определенный период времени.
KST_signal - KST	Пересечение скользящей средней KST и KST дает сигнал на открытие длинной позиции.

BB High - BB Low	Разность между верхней и нижней линиями Боллинджера показывает расстояние между ними.
sMACD_d, sRSI, sEMA, sStochastic	Разница между значением индикатора и его сдвинутым на единицу вперед значением. Дает информацию о направлении движения линии графика.

Таблица 1: Признаки

Разметка данных

Торговая система производит биржевые заявки: Short или Long. Для того, чтобы ограничить потери и зафиксировать прибыль используются Stop Loss и Take Profit. Большинство прогностических систем основано на том, что предсказывается движение цены на следующий день. В системе предсказывается, какая сделка (Short или Long) принесет доход с выставленными на определенном уровне Stop Loss и Take Profit. Разберем этот подход более подробно. Проведем итерацию по временному ряду OHLC (Row), где t перебирается от 1 до $\text{len}(\text{Row})-1$.

1. На шаге t фиксируем OpenPrice - цену открытия $t+1$ элемента ряда.
2. Фиксируем Stop Loss (SL) и Take Profit (TP) следующим образом:
 - Long Stop Loss = $\min(\text{EMA}(4), \text{OpenPrice}) - S * \text{ATR}(14)$
 - Short Stop Loss = $\max(\text{EMA}(4), \text{OpenPrice}) + S * \text{ATR}(14)$
 - Long Take Profit = $\max(\text{EMA}(4), \text{OpenPrice}) + P * \text{ATR}(14)$
 - Short Take Profit = $\min(\text{EMA}(4), \text{OpenPrice}) - P * \text{ATR}(14)$
3. Для $t+i$ (где i от $t+1$ до $\text{len}(\text{Row})$):
 - Если цена пробивает Long SL, больше не ждем, что сработает Long Take Profit.
 - Если цена пробивает Short SL, больше не ждем, что сработает Short Take Profit.
 - Если цена пробивает Long TP, присваиваем элементу ряда t метку 1.

- Если цена пробивает Short TP, присваиваем элементу ряда t метку -1.
- В случае, если не было присвоено метки 1 или -1, присваиваем метку 0.

4. Сдвигаем t : $t=t+1$.

В данном алгоритме S и P являются гиперпараметрами торговой системы.

Торговая система

В ходе обучения классификатора было замечено, что при многоклассовой классификации класс 0 в большинстве случаев не распознается. Поэтому из обучающей выборки мы удаляем данные с классом 0. Или же во время разметки данных ставим коэффициенты S и P достаточно большими, чтобы SL и TP не успели сработать. В таком случае 0 будут давать фильтры, рассмотренные ниже.

Другими словами, использование двухклассового классификатора в совокупности с определенными эвристиками для запрета торгов дают лучший результат, чем многоклассовая классификация, где класс 0 сообщает системе о необходимости пропустить бар и перестать торговать.

Модель, обученная таким образом, предсказывает направление позиции, т.е. метку класса -1 или +1. Эвристиками выступают два фильтра, запрещающие торговать.

Системе запрещено торговать, если:

- ADX меньше 20
- RSI меньше 10 или больше 90

Это позволяет уменьшить количество ошибок в случае отсутствия тренда и в случаях перекупленности/перепроданности рынка.

Кроме того, система отвергает ответ классификатора, если предсказанная вероятность принадлежности к классу меньше гиперпараметра T .

Тестирование

Тестирование производилось на фондовом индексе S&P 500. Гиперпараметры системы: S (коэфф. Stop Loss), P (коэфф. Take Profit), T

(порог вероятности принадлежности к классу). Sh - коэфф. Шарпа, Kl - коэфф. Кальмара, Profit - доход %, Failure - количество убыточных сделок, Success - количество прибыльных сделок. Все тесты проводились со стартовым капиталом - 200000 контрактов. Количество покупаемых/продаваемых контрактов - 50. Для сравнения в таблицу помещена стратегия Buy & Hold, рассчитанная для 1 контракта.

Таблица 2: Тестирование

№	S	P	T	Коэфф. Шарпа	Коэфф. Кальмара	Макс. Просадка %	Доход %	Success	Failure
1	3	0.5	0.52	-0.002	1.13	15	17	1190	170
2	4	0.6	0.53	0.026	5.07	14	71	1162	158
3	5	0.6	0.52	0.046	7.56	16	121	1362	140
-	-	-	-	0.018	2.21	57	126	1	0

Коэффициент Шарпа - это отношение средней премии за риск к среднему отклонению портфеля. Чем выше коэффициент, тем больше доходность портфеля компенсирует риски. Чтобы вычислить коэфф. Кальмара, требуется доходность поделить на максимальную просадку. Чем ниже коэффициент, тем выше риски для торговли.

Несмотря на то, что стратегия Buy & Hold имеет большую доходность, разработанная торговая система показала большую стабильность, т.к. в 2-х из 3-х случаев коэффициенты Шарпа и Кальмара сильно больше.

На графике 1 видно, что нет сильных просадок, и стоимость портфеля стабильно растет. Из этого можно сделать вывод, что модель действительно дает правильные направления движения инструмента. Также видно, что резкие развороты графика и периоды незначительного изменения цены (но большой волатильности) дают небольшую просадку стоимости портфеля.

Заключение

В данной работе была представлена торговая система, которая на основе градиентного бустинга и нескольких фильтров может генерировать сигналы на покупку или продажу. Для разметки данных был

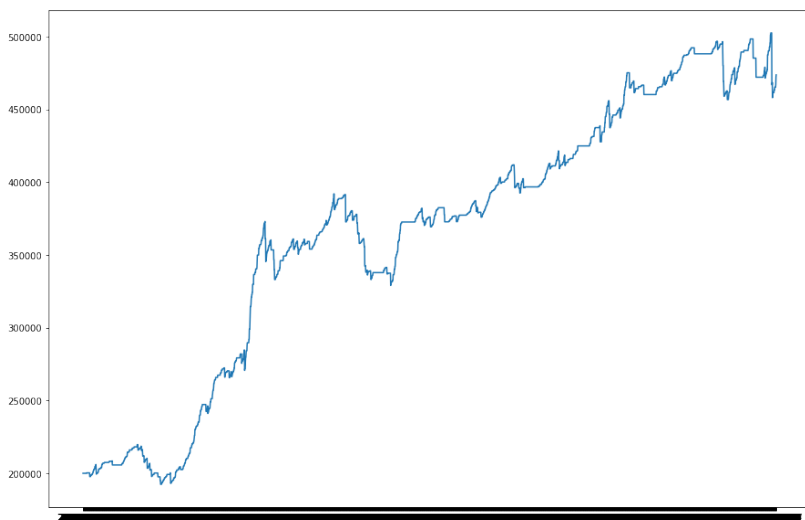


Рис. 1: График изменения цены портфеля для эксп. № 3

применен подход, основанный на Stop Loss и Take Profit. Система является достаточно гибкой, имеет настраиваемые гиперпараметры. Для дальнейшей модификации можно рассматривать возможность изменения состава индикаторов и автоматическое определение оптимальных значений параметров S , P и T .

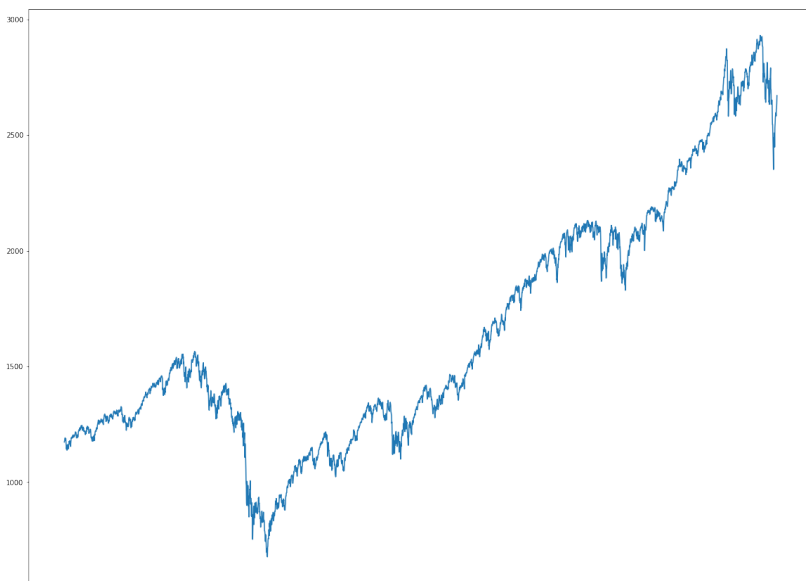


Рис. 2: График S&P 500 - Buy & Hold