

Инструкция по использованию

Классификатор Тренда v.1.0

Содержание

1. Установка системы на Windows
2. Описание предсказывающих алгоритмов
3. Описание сигналов системы
4. Описание параметров системы
5. Советы по использованию
6. Условия использования

Установка системы на Windows

Для правильного функционирования системы требуется скачать (и установить) следующие файлы:

- Терминал «MetaTrader 5»
- Терминал «R» (URL сайта для скачивания: <https://cran.r-project.org/bin/windows/base>)
- Файл индикатора «Classifier.ex5» (Поместить в папку, например: C:\MetaTrader 5\MQL5\Indicators)
- Файлы статистических функций «Space.func», «Space_2.func» (Создать папку «APS» в основной папке терминала, например: C:\MetaTrader 5\APS и поместить туда данные файлы)
- Файлы библиотеки, связывающей «MetaTrader 5» и терминал «R» mt4R.dll и mt4R.mqh (Файл mt4R.dll в папку, например: C:\MetaTrader 5\MQL5\Libraries, файл mt4R.mqh в папку, например: C:\MetaTrader 5\MQL5\Include)

Описание предсказывающих алгоритмов

Классификация алгоритмов машинного обучения, использованных в системе:

1. По назначению

- Предсказывающие направление тренда – основной сигнал (вероятность вверх или вниз)
- Предсказывающие качество основного сигнала (вероятность - высокая, неопределённость, низкая)
- Предсказывающие направление тренда на основе ансамбля алгоритмов (model stacking)

2. По типу

- Extreme Gradient Boosting (xgboost). Подробное описание <https://xgboost.readthedocs.io/en/latest/>
- Random Forest (rf). Подробное описание https://ru.wikipedia.org/wiki/Random_forest
- C5.0 algorithm (c5). Подробное описание <https://ru.wikipedia.org/wiki/C4.5>
- DeepBoost algorithm (deepboost). Подробное описание <https://research.google.com/pubs/pub42856.html>
- Multivariate Adaptive Regression Spline (earth). Подробное описание https://en.wikipedia.org/wiki/Multivariate_adaptive_regression_splines

Описание сигналов системы

В текущей версии системы реализовано 2 группы сигналов:

1. Индикация направления тренда

На гистограмме индицируется вероятность восходящего тренда, например, прогноз = 0,8 означает, что с вероятностью 80% цена будет двигаться вверх, и с вероятностью 20% ($1 - 0,8$) будет двигаться вниз, значение же 0, будет означать, что с вероятностью 0% тренд будет направлен вверх и с вероятностью 100% тренд будет направлен вниз.

Для удобства индикации за вероятность 100% принято значение 1 на гистограмме, а за вероятность 0% принято значение -1 на гистограмме.

2. Индикация степень достоверности текущего сигнала

На гистограмме индицируется вероятность правильного прогноза.

Степени достоверности:

- Прогноз правильный с большой вероятностью (показания: $[0.8 : 1]$, цвета Green или Red)
- Прогноз правильный с низкой вероятностью (показания: $[0.5 : 0.8]$, цвета LightGreen или Salmon)
- Прогноз не правильный с низкой или высокой вероятностью (показания $[0 : 0.5]$, цвет White)

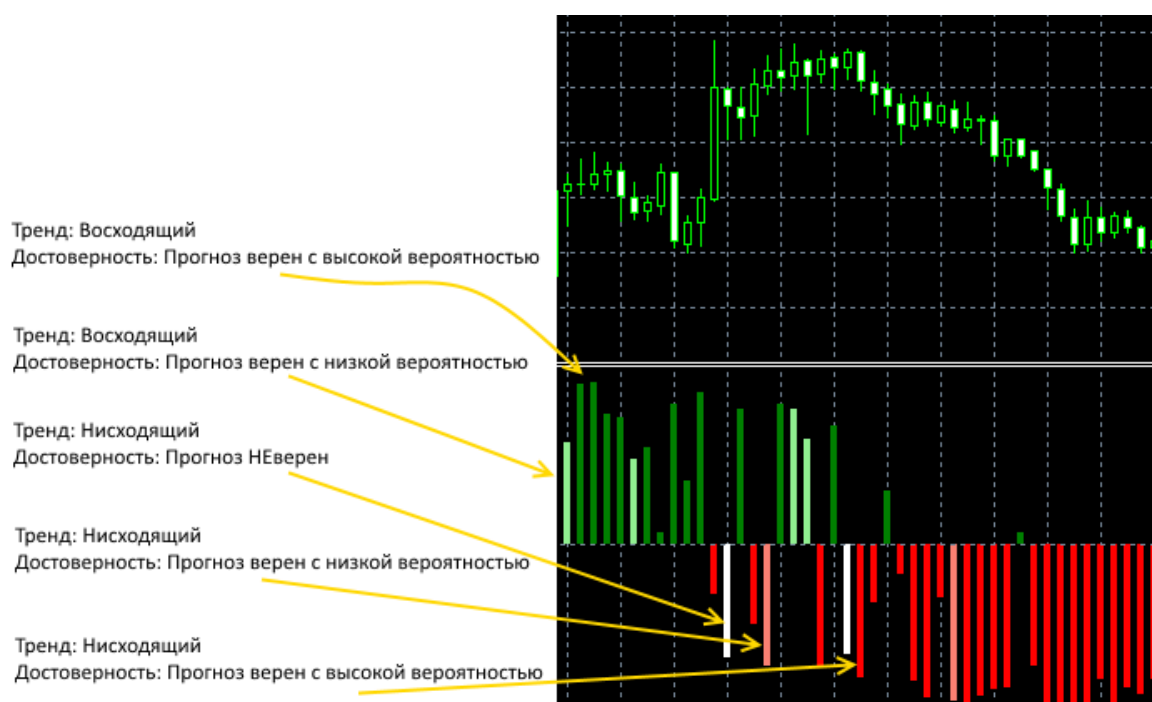


Рисунок 1 – Пример прогноза на 3 бара вперёд

Описание параметров системы

В данной задаче «Прогноз» - вероятность направления тренда вверх на отдельно взятом баре.

Признак (Feature) – переменная, помогающая более детально описать природу данных алгоритму.

Группа параметров «Основные»

rPath – Путь к файлу терминала R (Rterm.exe). Значение по умолчанию: "C:\Program Files\R\R-3.3.1\bin\i386\Rterm.exe".

predictTMF – Таймфрейм подачи данных для предсказания.

trainTMF – Таймфрейм подачи данных для обучения алгоритмов.

ZZ_DEPTH – Параметр индикатора ZigZag, отвечающий за определение длительности трендов, на которых будут обучаться алгоритмы. Чем больше данный параметр, тем более глобальные (длинные) тренды будут размечены в обучающей выборке. Значение по умолчанию: 10. Рекомендуемый диапазон значений: 3 – 150.

forecastStep – Горизонт прогнозирования, иными словами, на сколько баров вперёд будет предсказывать алгоритм. Если горизонт будет выставлен 1, то будет предсказываться тренд на следующем от строящегося бара. Если значение горизонта больше одного, то прогноз будет осуществляться для заданного количества баров и усредняться, тем самым можно получить средний прогноз на ближайшее “forecastStep” количество баров. Значение по умолчанию: 1. Рекомендуемый диапазон значений: 1 – (50% от LengthLearn).

LengthLearn – Длина обучающей выборки в барах. Чем больше данный параметр, тем больше алгоритм будет стремиться к обобщению в прогнозах (придавать меньшее значение отдельным колебаниям в истории), чем меньше, тем актуальнее прогнозы он сможет совершать (придавать значение только последним колебаниям). Время обучения прямо пропорционально длине обучающей выборки. Значение по умолчанию: 1000. Рекомендуемый диапазон значений: 200 – ∞ .

reTrainPeriod – Период обучения алгоритмов в барах. Используется для исполнения системы обучения «Скользящее окно» (актуализация данных алгоритмов для поддержания точности прогнозов на исходном уровне). Обычно

устанавливается в процентном соотношении 10 – 50% от длины обучающей выборки. Значение по умолчанию: 100. Рекомендуемый диапазон значений: 20 – ∞ .

reDrawPeriod – Период рисования прогнозных баров, служит для задания интервалов отрисовывания значений индикатора. Чем больше данный параметр, тем реже рисуются прогнозные бары, уменьшая время тестирования. Значение по умолчанию: 1. Рекомендуемый диапазон значений: 1 – ∞ .

Группа параметров «Алгоритмы прогнозирования».

В системе реализовано 5 кластеров для комбинирования разных алгоритмов. Значение каждого кластера может быть одним из видов моделей или «none», что означает не использование данного кластера.

typeOfMean – способ усреднения прогнозов от нескольких алгоритмов, может быть одним из следующих значений:

- arithmetic_mean – Среднее арифметическое прогнозов;
- geometric_mean – Среднее геометрическое прогнозов;
- stacking_of_models – Использование технологии «Стекинг алгоритмов» (Отдельно обучается алгоритм, который на вход принимает прогнозы от основных алгоритмов, а на выход подаёт обобщённый с помощью алгоритма XGboost прогноз);
- without_meaning – Без усреднения, устанавливается в том случае, если для прогнозирования будет использован только 1 алгоритм.

m1 ... m5 – Тип алгоритма для каждого кластера, может быть одним из следующих значений:

- xgboost – Алгоритм Extreme Gradient Boosting;
- xgboost_with_optimize – алгоритм Extreme Gradient Boosting с оптимизацией внутренних параметров самого алгоритма (сохраняются в папку C:\MetaTrader 5\APS\Optimize) и возможностью дальнейшей загрузки оптимизированных параметров, если в папке есть файл с оптимизированными параметрами для конкретного алгоритма, в ином случае алгоритм будет использован с внутренними параметрами по умолчанию;
- rf/rf_with_optimize – Алгоритм Random Forest, принцип оптимизации аналогичен xgboost;

- c5/c5_with_optimize – Алгоритм C 5.0, принцип оптимизации аналогичен xgboost;
- deepboost/deepboost_with_optimize – Алгоритм DeepBoost, принцип оптимизации аналогичен xgboost;
- earth,earth_with_optimize – Алгоритм Multivariate Adaptive Regression Spline, принцип оптимизации аналогичен xgboost;
- none – Кластер не используется для обучения алгоритма.

В машинном обучении, на первом месте стоят данные, поэтому следующие 2 группы параметров посвящены максимально полному описанию текущего состояния рынка, очистке и подготовке данных. Правильно подобранная и подготовленная комбинация признаков составляет 70% от создания обучающего алгоритма.

Группа параметров «Добавление дополнительных признаков к данным»

useDeepLearnFeatures – Использование признаков, созданных с помощью алгоритмов «Deep Learning» на исходных данных.

useTechIndicatorsFeatures – Использование признаков, созданных с помощью широко известных технических индикаторов.

useLastValAsFeatures – Использование предыдущих 10 наблюдений как признак. В алгоритмах классификации одному наблюдению (цены) соответствует одно значение целевой переменной (тренда), ввиду того, что используемые данные – это временной ряд, высока вероятность того, что предыдущие значения также влияют на будущее. В дальнейших версиях мы реализуем пользовательскую настройку количества таких значений, сейчас это 10 предыдущих наблюдений.

Группа параметров «Подготовка и очистка данных»

useNormalization – Использование нормализации данных. Позволяет снизить размерность данных и тем самым повысить точность прогнозирования. Для нормализации используется алгоритм «Метод пространственных знаков».

useDifferentiation – Использование ряда первых разностей цены. Позволяет избавиться от абсолютных значений, снизить размерность и тем самым повысить точность прогнозирования.

useBalanceOfClasses – Использование балансировки классов на обучающей выборке. Большинство алгоритмов машинного обучения очень чувствительны к выборкам, в

которых, наблюдения одного класса (тренд вверх) сильно превышают в количестве другой класс (тренд вниз), чтобы избавиться от этого можно уравнивать количество обоих классов.

`useHighCorrelationFeaturesCut` – Использование очистки высоко скоррелированных признаков. Нередко такие признаки не приносят никакой новой информации в алгоритм и являются избыточными, «зашумляя» основной сигнал и уменьшая точность.

`useNearZeroFeaturesCut` – Использование очистки низко дисперсных (редко или вообще не меняющихся во времени) признаков. Нередко такие признаки не приносят никакой новой информации в алгоритм и являются избыточными, «зашумляя» основной сигнал и уменьшая точность.

`useLinearCombosFeaturesCut` – Использование очистки линейно зависимых признаков. Нередко такие признаки не приносят никакой новой информации в алгоритм и являются избыточными, «зашумляя» основной сигнал и уменьшая точность.

Советы по использованию

Данная система разрабатывалась, как основной индикатор текущего тренда. Сейчас Вы можете определить направление текущего тренда или тренда через заданное количество баров.

Если вы запускаете индикатор в тестере, то таймфрейм тестирования должен соответствовать минимальному из двух значений – `predictTMF` и `trainTMF`. То есть, если параметр `trainTMF` = N4, а `predictTMF` = M5, то выставляется наименьший из 2х таймфреймов – M5.

Количество доступных данных должно быть достаточным для обучающей выборки. Если вы используете фьючерсные контракты для тестирования, то запускать индикатор лучше на «склейке».

Параметр «`forecastStep`» позволяет прогнозировать направление тренда на несколько баров вперёд. Это реализовано за счёт сдвига целевой переменной, относительно данных в будущее в обучающей выборке. Значение, которое будет возвращаться и отрисовываться индикатором является усреднением прогнозов на всю длину «`forecastStep`».

Большая обучающая выборка – не всегда приводит к лучшим результатам, т.к. финансовые данные это нестационарные временные ряды и для алгоритмов машинного обучения они плохо распознаваемы, но если взять такой промежуток времени, на которых данные были максимально приближены к стационарному ряду, то можно заметно повысить точность, а короткую обучающую выборку компенсировать небольшим периодом переобучения алгоритмов.

Обычно ансамбль из нескольких алгоритмов точнее прогнозирует, чем 1 отдельно взятый алгоритм, но данная гипотеза далеко не всегда верна, ввиду природы данных или особенностей конкретного алгоритма. Настоятельно рекомендуется начать тесты с 1го алгоритма и только потом, при необходимости, переходить на несколько. В дальнейшем планируется расширение количества алгоритмов прогнозирования.

Добавление большого количества признаков в алгоритм не всегда хорошо влияет на точность, т.к. есть вероятность, что признаки будут «зашумлять» основной сигнал (тренд). Настоятельно рекомендуется начать тесты без использования дополнительных признаков.

Условия использования

Продукт представлен пробной бета-версией. В дальнейшем планируется расширять функционал для повышения точности прогнозов.

Если вы нашли какие-либо ошибки в работе или хотите предложить идею для функционала, то пишите на почту automatps@gmail.com.

Ожидаемый результат совместного тестирования – торговая система, построенная на сигналах индикатора, если у вас есть идеи для совместной реализации, вы также можете написать на automatps@gmail.com.