

# Внедрение знаний И-Цзин в задачу прогнозирования с помощью интеллектуального анализа данных

Вэньцзе Лю, Нанкинский университет почты и телекоммуникаций,

Китай Сай Чэнь, Нанкинский университет почты и телекоммуникаций, Китай

Гояо Хуан, Нанкинский университет почты и телекоммуникаций, Китай

Линфэн Лу, Нанкинский университет почты и телекоммуникаций, Китай

Хуакан Ли, Сианьский университет Цзяотун-Ливерпуль, Китай

Гоцзы Сунь, Нанкинский университет почты и телекоммуникаций, Китай\*

## АБСТРАКТНЫЙ

Многие реальные приложения требуют прогнозирования, которое максимально использует преимущества данных. Классические механизмы интеллектуального анализа данных, как правило, снабжают модель прогнозирования основными данными для достижения многообещающего результата, который необходимо корректировать в различных сценариях применения. Недавние исследования показали потенциал механизма И Цзин для улучшения способности прогнозирования. Однако механизм прогнозирования И Цзин имеет несколько проблем, включая недостаточное использование знаний И Цзин и неполное преобразование данных. Чтобы решить эти проблемы, авторы разработали модель для использования знаний И Цзин и преобразования традиционной обработки прогнозов И Цзин в интеллектуальный анализ данных. Исследование авторов выявило многообещающие результаты на фондовом рынке по сравнению с популярными алгоритмами машинного обучения и глубокого обучения, такими как опорная векторная машина (SVM), экстремальный градиентный бустинг (XGBoost) и долгосрочная краткосрочная память (LSTM).

## КЛЮЧЕВЫЕ СЛОВА

Искусственный интеллект, интеллектуальный анализ данных, знание И-Цзин, машинное обучение, прогнозирование акций

## ВВЕДЕНИЕ

I Ching — это древний китайский документ, содержащий широкий спектр правил. Гадание по I Ching использует Инь и Ян и Четыре Знака для объяснения законов изменений в мире и предсказания всех видов дел. В настоящее время исследователи используют модель числовой гексаграммы I Ching для прогнозирования фондового рынка (Guo & Lu, 2020) и применяют гексаграммы для бизнес-руководства (Chen, 2021). Однако предыдущие механизмы прогнозирования акций на основе I Ching имеют несколько слабых мест: данные о небольшом масштабе фондового рынка приводят к неубедительным статистическим результатам, чрезмерно упрощенные показатели приводят к неполной начальной характеристике, а многие абстрактные концепции I Ching не могут быть преобразованы в доступный модус.

DOI: 10.4018/JDM.322097

\*Автор-корреспондент

Данная статья опубликована в открытом доступе и распространяется на условиях лицензии Creative Commons Attribution (<http://creativecommons.org/licenses/by/4.0/>) что разрешает неограниченное использование, распространение и воспроизведение на любом носителе при условии надлежащего указания автора оригинальной работы и источника оригинальной публикации.

Для решения вышеуказанных проблем авторы объединили механизм предсказания И Цзин и машинное обучение, представив модель предсказания на основе И Цзин. Традиционные методы предсказания И Цзин в основном используют объяснения Яо и гексаграммы, чтобы направлять людей к принятию разумных суждений.

Выбор признаков является важным шагом в построении модели прогнозирования И Цзин. Алгоритм выбора может отфильтровывать признаки с более высокой важностью из набора признаков высокой размерности. В настоящее время многие методы выбора признаков широко используются в оптической науке (Хуан и Лю, 2021) и области медицинской науки (Лю и др., 2019). Классические интегрированные алгоритмы выбора признаков, такие как случайный лес (RF) (Брейман, 2001), интегрируют результаты методов выбора признаков путем обучения нескольких признаков. В реальных приложениях три жизненно важных показателя особенно отражают влияние нечеловеческих факторов, таких как национальная политика и условия окружающей среды.

**Примечание.** Слева: Генератор получает необработанные наборы признаков. Авторы используют Three Vitals и Six Yao для тонкой настройки алгоритма RF и техники скользящих окон, что улучшает способность прогнозирования и гибкость. Справа: Декодер получает гексаграмму в качестве входных данных и извлекает процессор гексаграмм и пояснитель из «Zhou Yi».

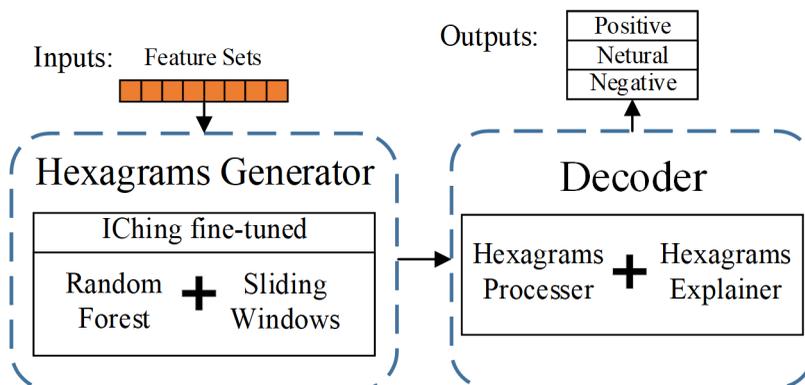
Мысль о Тэне, Чи и Цзинь в «Чжоу И» также называется «Три жизненными принципами», которые представляют небо, землю и человека. «Три жизненными принципами» утверждают, что люди являются отдельной сущностью, изолированной от неба и земли, но также признают, что человеческое поведение может влиять на работу всего мира. В частности, авторы предлагают метод выбора признаков, основанный на машинном обучении, чтобы в первую очередь ранжировать важность признаков, а затем использовать концепцию «Трех жизненных принципов», чтобы отфильтровать наиболее важные признаки в текущей сцене. Во-вторых, авторы используют алгоритм скользящих окон для преобразования признаков в гексаграмму; затем они определяют исходную гексаграмму и изменяют гексаграмму в соответствии с методом генерации гексаграммы И Цзин. Кроме того, авторы используют метод объяснения И Цзин для определения тренда акций и, наконец, получают результат прогнозирования, сравнивая прогнозируемое значение и значение тега. Объединение И Цзин и машинного обучения придает модели большую гибкость. Результаты эксперимента демонстрируют многообещающие результаты в задаче прогнозирования фондового рынка по сравнению с популярными алгоритмами машинного обучения и глубокого обучения, такими как машина опорных векторов (SVM), eXtreme Gradient Boosting (XGBoost), долговременная кратковременная память (LSTM) и рекуррентный блок вентилей (GRU).

## ФОН

Концепция Инь и Ян возникла из китайской культурной системы И Цзин и является основой древнекитайской философской мысли (Ли, 2014). Теория Инь и Ян показывает два разных состояния и учит людей видеть проблемы мира двумя способами. В реальных приложениях,

Рисунок 1.

Структура модели прогнозирования



Инь и Ян могут соответствовать как хорошим, так и плохим аспектам одной вещи. Любая сложная вещь может быть ограничена и разделена Инь и Ян.

Теория Четырех Знаков расширилась из концепции Инь и Ян (Ли, 2014). Это Старый Инь, Старый Ян, Молодой Инь и Молодой Ян, и они используются для отображения большего количества состояний вещей. В знании И Цзин вещи будут развиваться в противоположном направлении, когда они становятся экстремальными. Согласно этому, Старый Инь и Старый Ян изменятся на Молодой Ян и Молодой Инь, когда они находятся в экстремальных ситуациях. В реальной ситуации вещи не остаются в определенном состоянии все время. Например, акция на рынке ценных бумаг достигнет дна, когда упадет до определенного уровня, и цена начнет расти.

И Цзин в частности применяет путь природы к Трем Жизненным Силам неба, земли и людей. Люди живут в природе, с землей под ногами и небом над головой. Каждый Яос в гексаграммах соответствует различным местам в Трех Жизненных Силах.

Оригинальная гексаграмма представляет информацию, которая предсказывает начальную стадию вещей с помощью определенного метода, который фокусируется на прошлой и настоящей ситуации, в то время как измененная гексаграмма фокусируется на будущем. Гексаграммы не существуют независимо и должны быть объединены со словом и словом гексаграмм Яо для объяснения информации. Гексаграммы и слово Яо представляют собой текстовую последовательность, используемую для объяснения значения всех гексаграмм; есть 64 слова гексаграмм и 384 слова Яо. В частности, есть некоторые слова гексаграмм, показывающие опасное явление и побуждающие людей прекратить действие

или принять меры для уменьшения или избежания потерь, в то время как другие гексаграммы, показывающие

позитивное явление и побуждающее людей действовать смело и получать прибыль.

## СВЯЗАННЫЕ РАБОТЫ

В последние годы использование технологии интеллектуального анализа данных для макропрогнозов по данным в определенных сценариях постепенно стало популярной темой для исследований. Моделирование отфильтрованных данных и прогнозирование будущих тенденций по сути является проблемой многомерных временных рядов. Наиболее широко используемые многомерные модели временных рядов в основном включают в себя авторегрессионное интегрированное скользящее среднее (ARIMA) и LSTM. Прогнозируемые результаты ARIMA близки к историческому среднему и более подходят, когда есть больше шума, а реальные колебания значений относительно стабильны. LSTM достигает хороших результатов в многомерных временных рядах; однако, как модель глубокого обучения, она требует большого количества обучающих выборок, что приводит к строгим ограничениям для реальных приложений. ARIMA может достичь большей надежности с помощью алгоритмов онлайн-обучения (Liu et al., 2016), а LSTM может достичь лучшей производительности прогнозирования длинных последовательностей с помощью тензоризации. Когда механизм внимания добавляется поверх LSTM, он может сделать прогнозирование временных рядов более точным. Прогнозируемые результаты ARIMA близки к историческому среднему; он подходит, когда шум сильнее, а колебания реальных значений относительно стабильны.

Однако текущая модель макропрогнозирования не подходит для каждого сценария. Точность классических механизмов прогнозирования в задаче прогнозирования запасов едва превышает 50% (Kannan et al., 2010); оценка F1 RF и SVM в задаче прогнозирования спроса на велосипеды совместного пользования составляет до 62% и 51% по отдельности (Hu, 2013).

В процессе прогнозирования модели И Цзин, сначала авторы выбирают шесть признаков в качестве основы для шести Яо. Чтобы сделать выбор более научным и эффективным, авторы отфильтровывают 30 важных признаков, комбинируя алгоритм интеллектуального анализа данных. Существует много типов алгоритмов выбора признаков, включая фильтрующий выбор признаков, который может быстро удалять шумовые признаки, с высокой вычислительной эффективностью и сильной универсальностью. Эксперименты показывают значительные преимущества производительности при использовании наборов признаков, идентифицированных с точной взаимной информацией в задачах классификации (Брунато, 2016).

Алгоритм выбора обертывающих признаков также является типичным методом выбора признаков. Метод выбора обертывающих признаков использует отсутствующие данные для повышения производительности классификатора (Cao et al., 2016). Однако метод обертывания требует многократного обучения, вычислительные издержки велики по сравнению с методом фильтрации, и он не подходит для многомерных наборов данных.

Другой метод выбора признаков — это встроенный выбор признаков, который может обрабатывать многомерные наборы данных, но склонен к переобучению. Встроенный алгоритм выбора признаков, основанный

на вероятностной выходной чувствительности SVM для оценки важности конкретных признаков. Как типичный интегрированный алгоритм выбора признаков, RF (Breiman, 2001) широко используется в интеллектуальном анализе данных и других областях. RF имеет высокую точность прогнозирования и очень сильную толерантность к выбросам и шумным данным. Он обладает характеристиками, позволяющими давать множественные оценки важности при анализе многомерных данных. Эти преимущества делают RF очень подходящим для исследования многомерных данных, и он имеет высокую прикладную ценность в области интеллектуального анализа данных.

Рисунок 2 показывает, что Vital Features включают в себя две функции в каждой Vital с порядком Ten→Jin→Chi. Информация Yao в последовательности соответствует положению предыдущей функции. Затем необходимо восемь раз выполнить перестановку внутренней последовательности и сравнить каждую из них с упорядочением последовательности Yao. Процессор Hexagram просто следует правилам теории четырех знаков для преобразования YinYao и YangYao.

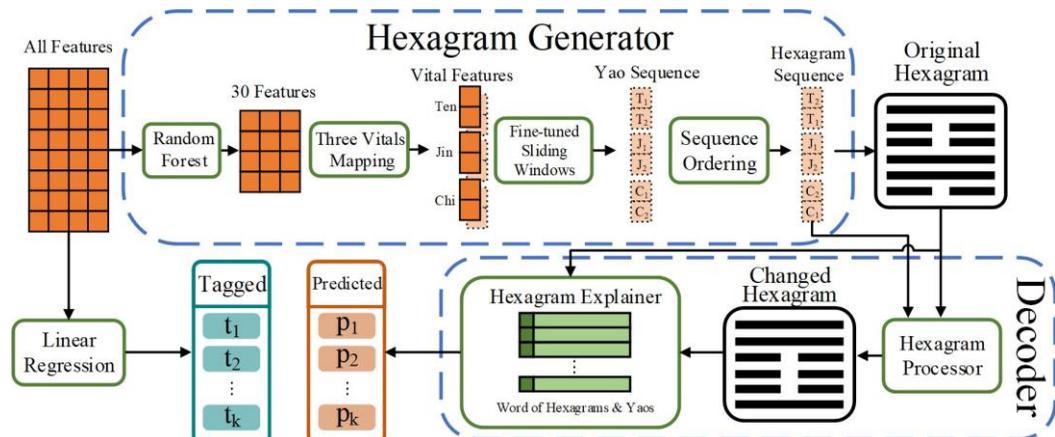
Чтобы расширить общность модели и улучшить производительность, авторы представляют макропрогнозирующую модель, которая объединяет технологию интеллектуального анализа данных и механизм прогнозирования И Цзин. И Цзин предсказывает будущие тенденции развития путем расшифровки гексаграмм. Механизм Трех Жизненных Факторов в И Цзин в сочетании с технологией интеллектуального анализа данных может отфильтровывать шесть важных признаков из большого количества признаков. Кроме того, через Четыре Знака в И Цзин Шесть Яо преобразуются в исходные гексаграммы, которые представляют собой 6-битные последовательности. После этого объяснитель гексаграмм генерирует измененные гексаграммы в соответствии с исходной. Наконец, система анализирует две последовательности и предоставляет результат прогнозирования через классификатор гексаграмм.

## МЕТОДОЛОГИЯ

Факторы, включенные в И Цзин, можно суммировать в фактор окружающей среды, географический фактор и человеческий фактор; их также называют тремя жизненными факторами, которые называются Тэн, Чи и Цзинь. В макроперспективе Тэн указывает на общее экологическое состояние вещей, на которое не будут влиять люди, Чи указывает на физические свойства и объективные законы, а Цзинь характеризует саму вещь. Эти макроскопические характеристики необходимо сочетать с фактической сценой при применении в реальном мире.

Применительно к фондовому рынку Ten соответствует национальной политике, мировой экономической ситуации и другим общим факторам окружающей среды, на которые не влияют компании, работающие с ценными бумагами, Chi соответствует размеру компании и данным о движении средств, а Jin соответствует фундаментальной информации компании (например, марже прибыли и коэффициенту задолженности).

Рисунок 2. Структура модели предсказания И Цзин



Другим важным свойством классической системы предсказаний И Цзин является Четыре Знака, которая воплощает динамическую природу И Цзин. В этой статье авторы используют метод скользящих окон для восстановления метода предсказания И Цзин. В частности, поскольку окно скользящее, значение Четырех Знаков, сформированное путем сопоставления данных в разные моменты времени, не является фиксированным. Классический метод предсказания И Цзин также рассматривается в декодере гексаграмм. Исследователи традиционного И Цзин используют различные методы для вывода изменения в гексаграммах, но по сути они используют биномиальное распределение в случайном процессе, который выполняется три раза. Поэтому начальный размер скользящих окон в модели предсказания И Цзин установлен равным восьми. Таким образом, авторы приводят современные алгоритмические методы в соответствие с традиционными методами предсказания И Цзин, а также улучшают способность модели соответствовать нелинейным отношениям.

## СТРОИТЕЛЬСТВО ЭТИКЕТКИ

В сценариях фондового рынка больше шумовых данных, и регрессия Хьюбера имеет хорошую надежность, что позволяет очень хорошо решать шумовые точки в данных. Авторы в основном используют регрессию Хьюбера для подгонки цены закрытия и получения тренда. В частности, используя одну линейную регрессию, выражение выглядит так:  $y=ax+b$ .  $y$  представляет прогнозируемое значение зависимой переменной,  $x$  представляет одну независимую переменную, а  $a$  и  $b$  являются неопределенными параметрами модели регрессии, где  $a$  также называется коэффициентом регрессии. Функция потерь регрессии Хьюбера — это потери Хьюбера, и она рассчитывается следующим образом:

$$L_{\delta}(y, \phi(x)) = \begin{cases} \frac{1}{2}(y - \phi(x))^2 & \text{для } |y - \phi(x)| \leq \delta \\ \frac{2}{3} \delta (y - \phi(x)) - \frac{1}{6} \delta^2 & \text{иначе} \end{cases} \quad (1)$$

где  $\delta$  — гиперпараметр, который необходимо скорректировать.

С помощью непрерывной итерации значение потерь минимизируется для получения оптимальной функции подгонки, и получается коэффициент регрессии  $a$  в функции. Согласно нормальному распределению, подсчитывается распределение  $a$ , и, наконец, категория метки получается в соответствии с интервалом распределения  $a$ .

## ВЫБОР ФУНКЦИЙ

В алгоритме RF можно сравнить вклад каждого признака в каждом дереве классификации и рассчитать индекс Джини, чтобы судить о важности каждого признака. Индекс Джини узла указывает на загрязненность узла. Вклад каждого индикатора риска в общий риск оценивается как процент среднего снижения Джини индикатора к сумме среднего снижения Джини всех индикаторов. Формула для расчета индекса Джини выглядит следующим образом:

$$GI_m = \sum_{k=1}^K p_{mk} (1 - p_{mk}) \quad (2)$$

$K$  — число категорий выборки,  $p_{mk}$  — вероятность того, что узел  $m$  принадлежит  $k$ -му образу, и, когда это бинарная классификация ( $K=2$ ), индекс Джини узла  $m$  равен:

$$GI = \frac{1 - \sum_{j=1}^M p_{ij}}{2} \quad (3)$$

вероятность того, что точка выборки принадлежит любой категории в узле  $m$ . Важность переменной  $X_j$  в узле  $m$  следующая:

$$VIM_{джм}^{(Джини)} = GI - GI - GI \quad (4)$$

где  $GI$  и  $GI$  соответственно представляют индексы Джини двух новых узлов, разделенных узлом  $m$ .

Если переменная  $X_j$  встречается  $M$  раз в дереве  $i$ , то важность переменной  $X_j$  в дереве  $i$  равна:

$$VIM_{ij}^{(Джини)} = \sum_{m=1}^M VIM_{джм}^{(Джини)} \quad (5)$$

Значимость Джини переменной  $X_j$  в РФ определяется следующим образом:

$$VIM_n^{(Джини)} = 1 - \dots^{(Джини)}$$

$$\text{джс} = \sum_{H=1}^n \text{ВИМ}_{ij} \quad (6)$$

где  $n$  — количество деревьев классификации в РФ.

## РАСЧЕТ ЧЕТЫРЕХ ЗНАКОВ И ГЕНЕРАЦИЯ ГЕКСАГРАММ

Сортируя важность каждой характеристики по индексу Джини этикетки, окончательно выбираются 30 основных характеристик с важностью индекса Джини. Необходимо сопоставить и классифицировать физическое значение характеристик акций с мыслями о трех жизненно важных показателях, а затем разделить все характеристики на три категории. После фильтрации 30 характеристик с помощью алгоритма RF, 30 характеристик могут быть сопоставлены с тремя категориями трех жизненно важных показателей в соответствии с таблицей соответствия между характеристиками и категориями трех жизненно важных показателей.

Наконец, необходимо выбрать две характеристики с наивысшей важностью в каждой категории, которая содержит в общей сложности шесть характеристик, и расположить отфильтрованные характеристики в порядке Чи, Цзинь и Тэн. После получения шести характеристик, согласно теории И Цзин, значения шести характеристик могут быть преобразованы в значения Четырех Знаков. Авторы используют механизм скользящих окон для отражения динамических характеристик И Цзин и вычисления значения Четырех Знаков.

Алгоритм первый: алгоритм генератора четырех знаков

Входные данные: размер окна  $k$ , отношение четырех знаков  $\text{old\_rate} = 1/8$ ,  $\text{young\_rate} = 3/8$ , список данных временного ряда  $\text{vec}=\{v_1, \dots, v_n\}$ , упорядоченный список  $\text{sortedvec}$  и нижний индекс элемента, добавленного в список раньше всех,  $m$ .

Вывод: список значений четырех знаков

1: для  $i = 0$  до  $k-1$  сделать

2:  $\text{inivec}[i] = \text{vec}[i]$

3: для  $j = k$  до  $n-1$

сделать

4:  $\text{sortedvec} = \text{sort}(\text{inivec})$

5:  $\text{maxv} = \max(\text{sortedvec})$ ,  $\text{minv} = \min(\text{sortedvec})$

6:  $\text{cha} = \text{maxv} - \text{minv}$

7:  $A = \text{старая\_ставка} * \text{ча}$ ,  $B = \text{молодая\_ставка} * \text{ча}$ .

8:  $\text{лаоинь} = [\text{минв}, \text{минв} + A]$ ,  $\text{шаоинь} = [\text{минв} + A, \text{мин} + A + B]$ ,

$\text{шаоян} = [\text{мин} + A + B, \text{мин} + A + 2B]$ ,  $\text{лаоян} = [\text{мин} + A + 2B, \text{максв}]$

9: если  $\text{vec}[j] \in \text{лаоинь}$

10:  $\text{sixiang\_list}[jk] = 6$

11: иначе если  $\text{vec}[j] \in \text{Шаоинь}$

12:  $\text{Sixiang\_list}[jk] = 8$

13: иначе если  $\text{vec}[j] \in \text{Шаоян}$

14:  $\text{Sixiang\_list}[jk] = 7$

15: иначе

16:  $\text{sixiang\_list}[jk] = 9$

17:  $\text{del sortedvec}[m], \text{sortedvec}[k] = \text{vec}[j]$

18: закончить, если

19: конец для

Гексаграмма состоит из Яо. Согласно идее И Цзин, ЯнЯо определяется как 1, а ИньЯо определяется как 0. Затем, сопоставляя значение Четырех Знаков с 1 и 0, значения 6 и 8 представляют ИньЯо, а 7 и 9 — ЯнЯо. С помощью механизма скользящих окон можно сгенерировать квартальную 6-битную последовательность для каждой акции. Исходная гексаграмма просто получается путем сортировки последовательности в порядке Чи, Цзинь и Тэн.

Алгоритм два: Объяснение гексаграмм

Ввод: Исходная гексаграмма *Orig<sub>i</sub>*, измененная гексаграмма *Change<sub>i</sub>*, измененное количество Яо *YbNum<sub>i</sub>*=0, 64 гексаграммы и словаря отображения объяснений *Gua<sub>-</sub>dict*, 384 Яо и словаря отображения объяснений

*Яо<sub>-</sub>дикт*, неизмененное Яо *Orig<sub>i</sub> Sta<sub>Y</sub>*, Яо, когда высокое положение не изменится у *Orig<sub>i</sub> Hp<sub>-</sub>Sta<sub>Y</sub>*, и Яо, когда низкое положение изменится у *Orig<sub>i</sub> Lp<sub>-</sub>Usta<sub>Y</sub>*.

Вывод: Пояснительный список исходных и измененных гексаграмм. 1: для

```

k = от 0 до 5 сделать
2: если ориг.я[k] != Изменитья[k]
3: флагя = 1, YbNumя = YbNumя+1
4: конец, если
5: конец для
6: если флагя = 0
7: резя = Gua-dicti[Origя]
8: конец, если
9: иначе
10: если YbNumя = 1
11: резя = Yao-dict[StaY]
12: иначе, если YbNumi = 2
13: резя = Яо-дикт[Hp-StaY]
14: иначе, если YbNumi = 3
15: резя = Gua-dict[Оригиналя]
16: иначе если YbNumя = 4
17: резя = Яо-дикт[Lp-UstaY]
18: иначе, если YbNumi = 5
19: резя = Yao-dict[StaY]
20: иначе
21: резя = Gua-dict[Изменитья]
22: конец, если

```

Согласно механизму И Цзин, гексаграммы изменяются, когда в последовательности Шести Яо будут Старый Ян Яо и Старый Инь Яо. Правило таково: каждая «1» в Старом Ян Яо изменится на «0», а каждый «0» в Старом Инь Яо изменится на «1». Наконец, выходные данные генератора гексаграмм определяются как исходная гексаграмма и измененная гексаграмма.

Декодер предназначен для определения категорий, соответствующих 64 гексаграммам в И Цзин и объяснений в каждой гексаграмме. Существует 64 различных объяснения гексаграмм и 384 объяснения Яо, которые строго связаны с каждой гексаграммой и положением Яо. Чтобы следовать оригинальному процессу объяснения, авторы извлекают алгоритм объяснения гексаграмм (Алгоритм Два) из классического механизма И Цзин.

Порядок Яос в Трех Жизненно важных аспектах неопределен и требует  $2*2*2=8$  перестановок, в общей сложности. Во время обучения авторы используют восемь различных результатов объяснения в качестве прогнозируемых значений и сравнивают их со значениями меток, соответственно. Наконец, авторы выбирают наиболее эффективный порядок Трех Жизненно важных аспектов для построения модели прогнозирования И Цзин в текущем сценарии.

Данные по фондовому рынку выбраны для проверки обоснованности этой модели, поскольку данные в финансовой сфере легко получить, а объем данных достаточно велик. Предложенная авторами модель применима и для данных в других областях. Ниже описаны этапы построения других моделей доменов.

Алгоритм третий: построение общей модели

Входные данные: Все функции *allfeature<sub>-</sub>list*, значение категории *tian<sub>-</sub>tian<sub>-</sub>label*, значение категории *di<sub>-</sub>di<sub>-</sub>label*, значение категории *class<sub>-</sub>ren<sub>-</sub>ren<sub>-</sub>label*, генератор четырех знаков *slideWin<sub>-</sub>algorithm*, правила формирования этой гексаграммы *guaXiang<sub>-</sub>o*, правила для

формирование изменения судьбы guaXiangc и гадание Интерпретация правил  
Hexagram\_Explain.

Параметр: tian\_list, di\_list, ren\_list, featurek, sixfeature\_list, labelSet

Вывод: Результаты предсказания гадания jiagua\_result.

```
1: для i = 0 до len(allfeature_list) сделать
2:   если allfeature_list[i] ∈ tian_label
3:     tian_list.append(allfeature_list[i])
4:   иначе если allfeature_list[i] ∈ di_label
5:     di_list.append(allfeature_list[i])
6:   еще
7:     ren_list.append(allfeature_list[i])
8:   конец, если
9:   конец для
10: линейнаяРегрессия(featurek) -> labelSet
11: randomForest(allfeature_list) и
   (tian_list, di_list, ren_list) -> sixfeature_list
12: fourSigns = slideWin_algorithm(sixfeature_list)
13: benGua = guaXiangc(fourSigns)
14: бянгуа = гуаянц (бэнгуа, четырезнака)
15: jiagua_result = Hexagram_Explain(benGua, bianGua)
16: сравнение между jiagua_result и меткой
17: получить результаты оценки модели
```

Вышеприведенный алгоритмический процесс поясняется следующим образом:

1. Все характеристики делятся на три категории, называемые тремя основными, а именно Ten, Chi и Jin. Ten соответствует характеристикам, которые могут отражать перспективы развития отрасли, Chi соответствует характеристикам, которые могут отражать текущее развитие рынка.

области, а Цзинь соответствуют характеристикам, которые могут отражать уровень развития самой области.

2. Необходимо выявить признаки, которые могут отражать тенденцию развития месторождения, и использовать алгоритм линейной регрессии для построения набора меток на основе этих признаков.
3. Алгоритм RF используется для расчета важности всех признаков, и выбираются два верхних признака в каждом Three Vital. В итоге получается шесть признаков.
4. Данные, соответствующие выбранным шести признакам, преобразуются в значения четырех признаков методом скользящего окна.
5. Согласно принципу создания гексаграмм в «И-Цзин», изначальная гексаграмма создается посредством значений Четырех Знаков.
6. Согласно принципу изменения гексаграмм в И-Цзин, измененные гексаграммы генерируются из исходных гексаграмм, а результаты предсказания получаются путем всестороннего суждения об исходных гексаграммах и изменяющихся гексаграммах. Наконец, точность предсказания достигается путем сравнения результата предсказания со значением метки.

## ЭКСПЕРИМЕНТЫ

### Экспериментальный набор данных

Необходимо настроить конкретный сценарий прогнозирования для проверки производительности модели прогнозирования И Цзин в реальном приложении. Необходимо учитывать следующие моменты: открытые наборы данных с легким доступом, крупномасштабные и простые в обслуживании и обновлении наборы данных, а также сценарий, насколько это возможно, образный. Таким образом, авторы в конечном итоге выбирают фондовый рынок в качестве экспериментального сценария. Тенденции акций делятся на три категории: падение, стабильность и рост. Линейная регрессия строит линейное представление, находя связь между независимыми и зависимыми переменными, которые часто используются для задач прогнозирования. Оптимизированный алгоритм линейной регрессии может значительно сократить ошибки прогнозирования и работать даже лучше, чем несколько алгоритмов глубокого обучения в определенном сценарии. Поэтому авторы также строят набор меток с помощью линейной регрессии. В частности, они подгоняют тенденцию цен закрытия за определенный период времени с помощью линейной регрессии и получают коэффициент регрессии. Наборы меток строятся из категорий в соответствии с диапазоном коэффициентов регрессии.

Наборы данных фондового рынка, которые авторы использовали в эксперименте, были взяты из NetEase Finance.Веб-сайт, который собрал данные по 3000 различных акций с марта 2010 года по март 2020 года. Наборы данных состоят из двух частей, а именно финансовых данных и данных о движении денег. В частности, финансовые данные сообщают о чистой прибыли компании, коэффициенте задолженности и других данных, отражающих результаты компании в каждом квартале, а данные о движении денег сообщают о цене открытия компании, цене закрытия, скорости оборота и других данных о движении капитала ежедневно. Все акции разделены на 10 секторов в соответствии с отраслью, и эксперимент предсказывает тренд акций в следующем году на основе исторических данных за прошлый квартал.

### Индикаторы оценки

Обычно используемыми показателями оценки для задач классификации являются точность, полнота и оценка F1. В таблице 1 показана матрица путаницы.

Точность относится к доле всех правильных прогнозов в общем количестве, и формула выглядит следующим образом:

Таблица 1.  
Матрица  
путаницы

	Положительный	Отрицательно
Истинный	Истинный положительный	Истинный отрицательный
ЛОЖЬ	Ложный положительный результат	Ложный отрицательный результат

$$P = \frac{ТП}{ТП + ФП} \quad (7)$$

Полнота относится к доле правильных положительных предсказаний ко всем правильным предсказаниям. Формула выглядит следующим образом:

$$R = \frac{ТП}{ТП + ФН} \quad (8)$$

Оценка F1 — это показатель, объединяющий P и R. Формула выглядит следующим образом:

$$F1 = \frac{2PR}{P+R} = \frac{2ТП}{2ТП + ФП + ФН} \quad (9)$$

Оценка F1 позволяет всесторонне оценить производительность классификатора. Чтобы проверить возможности модели, авторы в основном провели две группы сравнительных экспериментов: (1) они сравнили модель прогнозирования И Цзин с другими классическими алгоритмами с учетом и без учета отраслевых различий; (2) они проверили влияние отраслевых различий на модель прогнозирования И Цзин.

## ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ И АНАЛИЗ

Авторы выбрали классические алгоритмы машинного обучения для сравнительных экспериментов, включая SVM, XGBoost и k-ближайший сосед (KNN). XGBoost — это реализация алгоритма бустинга (Tianqi, Chen et al., 2016).

Рисунок 3 показывает, что модель И Цзин работает лучше, чем другие классические алгоритмы в задаче прогнозирования фондового рынка (рисунок 3а), точность увеличивается с длиной прогноза и сходится, когда длина прогноза достигает 12 месяцев (рисунок 3б), и, в отличие от других алгоритмов, модель прогнозирования И Цзин показывает значительное улучшение производительности в моделировании отрасли, а Id.x предоставляет отрасль, упомянутую авторами выше, в том же порядке размерной проблемы катастрофы и склонную к переобучению и недообучению (рисунок 3в).

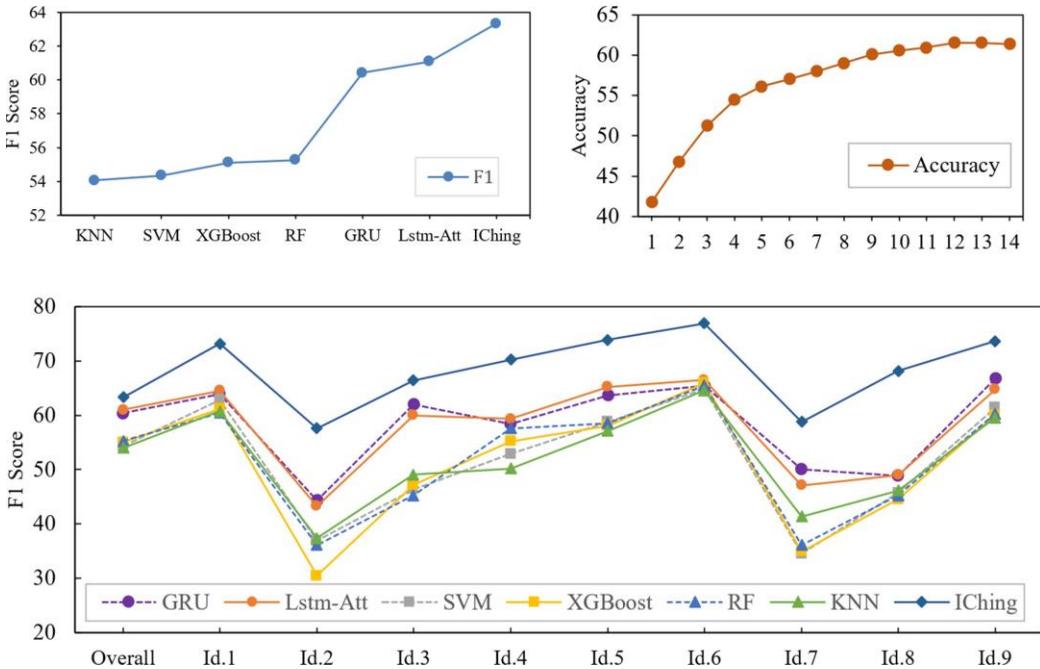
XGBoost состоит из нескольких базовых обучающихся, и каждая итерация основана на разнице между предыдущим значением и целевым значением. Предсказанное значение XGBoost состоит из всех предыдущих значений, что дает модели хорошую способность к антипереобучению. Еще одним преимуществом XGBoost являются параллельные вычисления, которые повысили скорость обучения.

SVM — это алгоритм классификации, ориентированный на данные (Cortes & Vapnik, 1995). Основан на линейных признаках обучающих данных, существует три типа модели обучения SVM: линейный SVM, линейно разделимый SVM и нелинейный SVM. Преимущества SVM в том, что он имеет хорошую производительность в небольших выборках и решает проблемы размерной катастрофы и вычислительной сложности. Однако SVM не подходит для решения многоклассовых задач.

KNN может использоваться для классификации или регрессии. KNN оценивает характеристики образца в соответствии с соседями образца в пространстве признаков. Преимущества KNN в том, что он прост в реализации, не требует обучения и особенно подходит для задач мультиклассификации. Тем не менее, KNN покажет очевидную ошибку, если образец не сбалансирован.

Внимание LSTM использует механизм внимания в LSTM и приводит к лучшим результатам. В реальном применении важность признаков долгосрочных рядов является мутационной и не может быть различена LSTM. Механизм внимания может решить эту проблему, взвешивая выход скрытого слоя LSTM. При построении модели слой внимания и полностью связанный слой должны быть добавлены после слоя LSTM.

**Рисунок 3.**  
Оценка F1 классических алгоритмов, точность увеличения длины прогноза и оценка F1 отраслевого моделирования



GRU — это тип рекуррентной нейронной сети, и предлагается для решения таких проблем, как долговременная память и градиенты в обратном распространении. По сравнению с LSTM, GRU использует шлюз обновления для управления объемом данных, которые предыдущая информация памяти может продолжать сохранять до текущего момента, и шлюз сброса для управления тем, сколько прошлой информации следует забыть.

Классические методы машинного обучения работают лучше, чем глубокое обучение, поскольку цена закрытия акций зависит от многих факторов, и модели нейронной сети сложно изучать данные со слабой регулярностью. С другой стороны, классические алгоритмы машинного обучения сложнее решить.

Производительность модели предсказания И Цзин выше, чем у других алгоритмов по всем оценкам Индикаторы, за исключением отзыва на фондовых наборах данных. Модель прогнозирования И Цзин может выбирать различные индикаторы в качестве основы для прогнозирования в соответствии с различными условиями и имеет хорошую адаптивность в различных сценариях.

**Таблица 2.**  
Сравнение эффективности модели на фондовом рынке с классическим алгоритмом

Алгоритм	Точность	Отзывать	Оценка F1
СВМ	45.51	67.46	54.35
XGBoost	46.31	68.05	55.11
РФ	46.46	68.16	55.26
ГРУ	60.15	60.71	60.43
LSTM-Att	61.85	60.33	61.08
КНН	52.00	56.77	54.05
И Цзин	65.41	61.51	63.33

Таблица 3.  
Экспериментальные результаты моделирования отрасли И Цзин

Промышленность	П	Р	Оценка F1
Общий	65.41	61.51	63.33
Транспортная логистика	78.81	68.65	73.20
Информационные технологии	56.64	58.46	57.53
Химикаты	70.14	63.36	66.43
Строительство	74.15	67.13	70.22
Недвижимость	80.49	68.87	73.99
Оптовая и розничная торговля	71.61	64.16	67.56
Сила и энергия	84,56	70,84	76,94
Устройство связи	58.99	58.45	58.72
Общее производство	71.80	65.09	68.17
Добыча полезных ископаемых	79.99	68.73	73,69

После моделирования по отраслям точность большинства отраслей выше, чем общее моделирование, и улучшение не стабильно. Учитывая, что характеристики каждой отрасли различны, модель И Цзин может строить гексаграммы на основе характеристик каждой отрасли и достигать лучшей производительности и гибкости прогнозирования. Однако две отрасли работают хуже, чем общее моделирование: информационные технологии и общее производство. Причина может заключаться в том, что параметр `win_len` не является оптимальным значением. Учитывая быстрое развитие в этих двух отраслях, развитие отрасли может кардинально измениться всего за несколько месяцев. Кроме того, как отрасль, которая сейчас привлекает к себе много внимания, она будет подвержена влиянию более неопределенных факторов, таких как национальная политика и ориентация обществу.

Преимущества метода, предложенного авторами в данной статье, по сравнению с традиционными алгоритмами интеллектуального анализа данных в основном включают два аспекта:

1. Модель И Цзин в значительной степени избегает риска переобучения, поскольку модель, построенная в этой статье, использует уменьшение размерности признаков, линейную и нелинейную подгонку для достижения признака функции активации. Сама модель несложная, и экспериментальные результаты могут показать, что производительность модели И Цзин на тестовом наборе лучше, чем у традиционной модели добычи данных при сравнительной проверке.
2. Традиционные модели добычи данных требуют крупномасштабных наборов данных для обучения и предъявляют высокие требования к качеству наборов данных. Метод авторов позволяет быстро выбирать ключевые признаки, используя преимущественно И Цзин. Модель И Цзин должна сосредоточиться только на выборе признаков Трех Жизненных Сущностей, значениях Четырех Знаков и измененном Яо, что требует меньше параметров, может быстро свести модель с меньшим количеством наборов данных и может получить лучшие результаты прогнозирования. Кроме того, модель И Цзин обладает сильной универсальностью. Когда она применяется к другим областям, ей нужно сосредоточиться только на построении картографических отношений между признаками и сущностями Трех Жизненных Сущностей в рамках фоновых знаний, и нет необходимости вносить коррективы в структуру модели.

## ПОДТВЕРЖДЕНИЕ

Это исследование было завершено под руководством профессора Гоцзы Суя и Хуа Канга Ли. Их профессиональные знания, научное отношение и строгий академический дух

оказали значительное влияние на эту работу.

## ЗАКЛЮЧЕНИЕ

В этой статье авторы исследовали задачу прогнозирования и предложили модель прогнозирования, основанную на механизме И Цзин, в частности, тонко настроенную модель прогнозирования И Цзин, RF и механизм скользящих окон для ранжирования всех признаков по важности. Модель И Цзин авторов включает классический процесс генерации гексаграмм И Цзин и декодирует гексаграммы в соответствии с оригинальным текстом Чжоу И, который глубоко интегрирует знания И Цзин. В частности, авторы использовали Три жизненно важных элемента для снижения размерности признаков. Процесс отображения Четырех Знаков похож на функцию, которая знает независимую переменную для получения зависимой переменной, поскольку динамические признаки являются ключевыми признаками И Цзин, а результат отображения Четырех Знаков является относительным, а не абсолютным. Поэтому авторы использовали динамический метод скользящего окна. Согласно концепции исходной гексаграммы, исходная гексаграмма формируется из значений Четырех Знаков, что завершает работу генератора. Кроме того, в И Цзин есть изменение гексаграмм, которое формируется на основе исходной гексаграммы по правилам изменения Яо. Процесс изменения гексаграммы в основном является линейным преобразованием по изученным признакам. Декодер завершает декодирование и получает результат предсказания после того, как модель получает исходные гексаграммы и измененные гексаграммы, а также изменяется количество и положение Яо.

Для проверки возможности применения в реальном мире авторы провели эксперимент в сценарии фондового рынка. Эксперименты показали, что модель прогнозирования И Цзин имеет хорошую производительность в задаче прогнозирования тенденций акций. Однако, с другой стороны, модель прогнозирования И Цзин не корректирует динамически скользящее окно из-за скорости развития и цикла роста различных отраслей.

Прогнозирование тренда — это только один аспект добычи ценности больших данных на фондовом рынке. В будущей работе исследование диагностики проблем и раннего предупреждения также может быть проведено на основе метода прогнозирования, который авторы проиллюстрировали в этой статье. Когда прогнозируется, что будущее развитие акций будет неблагоприятным, модель определяет причину неблагоприятной ситуации и направляет компанию на корректировку своей бизнес-стратегии в соответствии с причиной, чтобы достичь роли диагностики и раннего предупреждения.

Кроме того, авторы не использовали в эксперименте данные новостных текстов; новостные данные могут отражать тенденция развития отрасли, которая полезна для прогнозирования будущего развития компании. В последующей работе текстовые данные также должны быть добавлены для обучения.

## КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют, что у них нет никаких коммерческих или ассоциативных интересов, представляющих собой конфликт интересов, в связи с представленной работой.

## ФИНАНСОВОЕ АГЕНТСТВО

Данное исследование было поддержано Национальным фондом естественных наук Китая [номера грантов 61906099] и Открытым фондом Ключевой лаборатории мониторинга и моделирования городских земельных ресурсов Министерства природных ресурсов [номера грантов KF-2019-04-065].

## ССЫЛКИ

Audemard, G., Bellart, S., Bounia, L., Koriche, F., Lagniez, JM, & Marquis, P. (2022). Торговля сложностью ради разреженности в объяснениях случайного леса. Труды конференции AAAI по искусственному интеллекту. doi:10.1609/aaai.v36i5.20484

Bartley, C., Liu, W., & Reynolds, M. (2019). Улучшенные алгоритмы случайного леса для частично монотонной порядковой классификации. Труды конференции AAAI по искусственному интеллекту. doi:10.1609/aaai.v33i01.33013224

Бартол, К., Боянич, Д., Петкович, Т., Пехарец, С. и Прибанич, Т. (2022). Линейная регрессия против глубокого обучения: простая, но эффективная базовая линия для измерения человеческого тела. Датчики (Базель), 22(5), 1885. doi:10.3390/c22051885 PMID:35271032

Брейман. (2001). Случайные леса. Mach Learn.

Брунатом, Б. (2016). X-MIFS: Точная взаимная информация для выбора признаков. В трудах Международной совместной конференции по нейронным сетям 2016 года (IJCNN). IEEE. doi:10.1109/IJCNN.2016.7727644

- Сао, ТТ, Zhang, MJ, & Andreae, P. (2016). Подход к выбору признаков-оболочек для классификации с отсутствующими данными. Труды 19-й Европейской конференции по приложениям эволюционных вычислений.
- Чандрашекар, Гириш и Ферат. (2014). Обзор методов выбора признаков. Компьютеры и электротехника.
- Чэнь, ЙВ (2021). Как предприятия сталкиваются с трудностями? Из И Цзин просветление «Четырех трудных триграмм». Обзор менеджмента Цинхуа.
- Кортес, К. и Вапник, В. (1995). Сети опорных векторов. Машинное обучение, 20(3), 273–297. doi:10.1007/BF00994018
- Фан, ХН (2021). Введение в Априорную И Цзин: Феноменологическое исследование И Цзин. Исследования И Цзин.
- Гораде, СМ, Део, А. и Пурохит, П. (2017). Исследование некоторых методов классификации интеллектуального анализа данных. *Международный Научно-исследовательский журнал по инженерингу и технологиям*, 4(4), 3112–3115.
- Го, И. и Лу, С. (2020). Анализ применения числовой гексаграммной модели И-Цзин в прогнозировании фондового рынка. Современный маркетинг.
- Ху, И. (2013). Модель прогнозирования для фондового рынка: сравнение крупнейших инвесторов мира с методом интеллектуального анализа данных. WHISEB.
- Хуан, С. и Лю, В. П. (2021). Исследование выбора признаков в ближнем инфракрасном диапазоне на основе переменной важности и частичных наименьших квадратов. Журнал Университета города Хунань (естественные науки).
- Исмаил, Ф. (2019). Глубокое обучение для классификации временных рядов: обзор. *Data Mining and Knowledge Discovery*, 33(4), 917–963. doi:10.1007/s10618-019-00619-1
- Каннан, К. С., Секар, П. С. и Мохамед, С. М. (2010). Финансовый прогноз фондового рынка с использованием методов интеллектуального анализа данных. Конспект лекций по инженерингу и информатике.
- Кесаварадж, Г. и Шрикумар, С. (2013). Исследование методов классификации в интеллектуальном анализе данных. В трудах Четвертой международной конференции по вычислительным, коммуникационным и сетевым технологиям (ICCCNT) 2013 года. IEEE. doi:10.1109/ICCCNT.2013.6726842
- Ким, С. (2021). Прогнозирование стоимости строительства трубопровода с использованием методов многомерных временных рядов. Журнал Pipeline *Системная инженерия и практика*.
- Ким. (2011). Модель классификации типов потребностей клиентов с использованием методов интеллектуального анализа данных для рекомендательных систем. *Международный журнал экономики и управленческой инженерии*.
- Куй Ю., Сяньцзе Г., Линь Л., Цзююн Л., Хао В., Чжаолун Л. и Синдун В. (2020). Функция, основанная на причинно-следственной связи Выбор: Методы и оценки. *ACM Computing Surveys*, 53(5).
- Ли, СУ (2014). «Единство человека и природы» или «Тен, Чи, Цзинь» Три жизненных принципа: базовая структура конфуцианской философии окружающей среды. Исследования И Цзин.
- Liu, C., Noi, SC, Zhao, P., & Sun, J. (2016). Онлайн-алгоритмы ARIMA для прогнозирования временных рядов. Труды конференции AAAI по искусственному интеллекту. doi:10.1609/aaai.v30i1.10257
- Лю, УХ, Чэнь, Б. и Чжоу, ЗУ (2019). Улучшенный алгоритм выбора признаков на основе случайного леса. Современные электронные технологии.
- Макия, Н., Алекс, С., Ёнсу, К., Джонхён, К. и Джино, К. (2021). Автоматизированный выбор признаков для аномалии обнаружение в данных сетевого трафика. Труды ACM по системам управленческой информации, 12(3).
- Nascimento Da Silva, P., Plastino, A., Fabris, F., & Freitas, AA (2021). Новый метод выбора признаков для неопределенных признаков: применение к прогнозированию генов про-/анти-долголетия. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 18(6), 2230–2238. doi:10.1109/TCBB.2020.2988450 PMID:32324561
- Ниламегам, С. и Рамарадж, Э. (2013). Алгоритм классификации в интеллектуальном анализе данных: обзор. Сетевые тенденции и технологии: международный журнал.
- Рана, А. (2021). Обзор методов интеллектуального анализа данных для анализа больших данных. Азиатский журнал многомерных исследований.
- Ратанамахатана, К. А. и Имонн, К. (2005). Три мифа о динамическом искажении времени при добыче данных. В трудах Международной конференции SIAM 2005 года по добыче данных. Общество промышленной и прикладной математики. doi:10.1137/1.9781611972757.50
- Руис, А., Флинн, М., Лардж, Дж., Миддлхерст, М. и Бэгнолл, А. (2021). Великий многомерный эксперимент по классификации временных рядов: обзор и экспериментальная оценка последних достижений алгоритмов. *Data Mining and Knowledge Discovery*, 35(2), 401–449. doi:10.1007/s10618-020-

00727-3 PMID:33679210

Руис, А., Пасос, М., Флинн и Бэгнолл, А. (2020). Сравнительный анализ алгоритмов классификации многомерных временных рядов. Академическая пресса.

Shi, Q., Yin, J., Cai, J., Cichocki, A., Yokota, T., Chen, L., Yuan, M., & Zeng, J. (2020). Блочный ганкелев тензорный ARIMA для прогнозирования нескольких коротких временных рядов. Труды конференции AAAI по искусственному интеллекту. doi:10.1609/aaai.v34i04.6032

Сильвестрини, А. и Вередас, Д. (2008). Временное агрегирование одномерных и многомерных моделей временных рядов: обзор. Журнал экономических обзоров, 22(3), 458–497. doi:10.1111/j.1467-6419.2007.00538.x

Tang, J., Salem, A., & Huan, L. (2014). Выбор признаков для классификации: обзор. Классификация данных: алгоритмы и приложения.

Tianqi, C., & Guestrin, C. (2016). XGBoost: масштабируемая система повышения эффективности деревьев. Клиническая ортопедия и смежные исследования.

W, XN (2015). Составление слов И Цзин Гуа и Яо: триада литературы, истории и философии.

*Социальные науки Цзянси.*

Ван, В. и Каррейра-Перпинан, М. (2014). Роль снижения размерности в классификации. Труды конференции AAAI по искусственному интеллекту. doi:10.1609/aaai.v28i1.8975

Xu, D., Cheng, W., Zong, B., Song, D., Ni, J., Yu, W., Liu, Y., Chen, H., & Zhang, X. (2020). Тензоризированная LSTM с адаптивной общей памятью для изучения тенденций в многомерных временных рядах. Труды конференции AAAI об искусственном интеллекте. doi:10.1609/aaai.v34i02.5496

Y, H., & H, YD (2011). Обсуждение концепции «изменения» в «И Цзин». Chuanshan Academic Journal.

Юэ, З., Ван, И., Дуань, Дж., Ян, Т., Хуан, К., Тонг, И. и Сюй, Б. (2022). TS2Vec: На пути к универсальному представлению временных рядов. Труды конференции AAAI по искусственному интеллекту. doi:10.1609/aaai.v36i8.20881

Чжан, И. и Ю, ХН (2019). Исследование макроэкономических факторов, влияющих на стабильность фондового рынка. Труды Ежегодного собрания по управленческой инженерии 2019 г. doi:10.1145/3377672.3378037

Чжао, П. и Лай, Л. (2021). Эффективная классификация с адаптивным KNN. Труды конференции AAAI по искусственному интеллекту.

Zheng, HY, Zhou, ZQ, & Chen, JY (2021). RLSTM: Новая структура прогнозирования запасов с использованием случайного шума для предотвращения переобучения. Computational Intelligence and Neuroscience, 2021(3), 1–14. doi:10.1155/2021/8865816 PMID:34113377

Чжоу, З. и Хукер, Г. (2021). Непредвзятое измерение важности признаков в древовидных методах. ACM Transactions on Knowledge Discovery from Data, 15(2), 1–21. doi:10.1145/3429445

*Вэньцзе Лю — студент магистратуры Нанкинского университета почты и телекоммуникаций. Научные интересы Вэньцзезанимаются добычей данных и обработкой естественного языка.*

*Чэнь Сай — магистр Нанкинского университета почты и телекоммуникаций. Научный интерес Чэня — интеллектуальный анализ данных и обработка естественного языка.*

*Гояо Хуан — студент Нанкинского университета почты и телекоммуникаций.*

*Линфэн Лу — магистр Нанкинского университета почты и телекоммуникаций. Научный интерес Линфэна — обработка естественного языка.*

*Хуакан Ли получил степень доктора философии в области компьютерных наук в Школе компьютерных наук и инженерии Университета Айдзу в 2011 году. В настоящее время Хуакан Ли работает доцентом в Школе искусственного интеллекта и передовых вычислений, Предпринимательского колледжа ХJTLU (Тайцан). С мая 2011 года по сентябрь 2013 года Хуакан Ли работал научным сотрудником-постдоком на кафедре компьютерных наук и инженерии Шанхайского университета Цзяотун. В этот период Хуакан Ли отправился в Ali Cloud Computing Ltd. в качестве приглашенного исследователя для разработки Национального проекта 863. С сентября 2013 года по июль 2020 года Хуакан Ли был преподавателем в Школе компьютерных наук Нанкинского университета почты и телекоммуникаций. Его исследования сосредоточены на искусственном интеллекте и интеллектуальном анализе больших данных, в частности на обработке естественного языка, социальных вычислениях, инженерии знаний, графах знаний и связанных с ними приложениях в этой области.*

*Guozi Sun — профессор Школы компьютерных наук и технологий в Нанкинском университете почты и телекоммуникаций, Китай. Его исследовательские интересы включают криминалистику блокчейна, цифровую криминалистику и цифровые расследования. Sun получил докторскую степень в области машиностроения и автоматизации в Нанкинском университете авиации и астронавтики, Китай. Он является членом IEEE*

**Журнал управления базами  
данных**

*Computer Society, ACM, Китайской компьютерной федерации (CCF), Китайского института электроники (CIE) и Общества информационной безопасности и криминалистики (ISFS), Китай.*