

Грязнов Сергей Александрович

канд. пед. наук, доцент, декан

ФКОУ ВО «Самарский юридический институт ФСИН России»

г. Самара, Самарская область

ПРОГНОЗИРОВАНИЕ ИНФЛЯЦИИ С ПОМОЩЬЮ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ

Аннотация: статья посвящена вопросам прогнозирования инфляции в среде с большим объемом данных, что является сложной задачей и активной областью исследований. Один из методологических вопросов заключается в том, как найти и извлечь информацию с наибольшей предсказательной силой для интересующей переменной, если имеется много высококоррелированных предикторов, как в задаче прогнозирования инфляции. Традиционно для решения этой проблемы использовались факторные модели, однако сегодня модели машинного обучения, такие как случайные леса, могут предложить более эффективные решения проблемы.

Ключевые слова: прогнозирование инфляции, макроэкономические показатели, машинное обучение, случайный лес, моделирование многомерных данных, прогностический анализ.

Один из важнейших индикаторов состояния экономики – инфляция. Прогноз национальной инфляции важен не только для государственных структур, но и для бизнеса и домохозяйств – будь то решение об утверждении корпоративной стратегии или решение о плане крупных покупок. В глобальном масштабе задача повышения стабильности экономики обычно выполняется центральными банками. Сложность прогнозирования макроэкономических показателей обусловлена наличием множества сильно коррелированных переменных, которые влияют на прогнозируемую переменную. Для решения этой проблемы как правило используются факторные модели, благодаря которым, вместо помещения всех переменных в уравнение прогнозирования, моделируются только несколько факторов. Однако при этом неясно, даст ли использование всех име-

ющихся независимых переменных или их сокращенного набора с помощью алгоритмов предварительного отбора, более информативные извлеченные факторы. Поэтому факторные модели, основанные на большем количестве переменных, могут быть менее эффективны при прогнозировании по сравнению с моделями, полагающимися на небольшое количество переменных. Кроме того, еще одной нерешенной проблемой является выбор количества факторов, которые будут использоваться в уравнении прогнозирования. Альтернативный подход к прогнозированию экономической переменной в средах с большим объемом данных заключается в использовании преимуществ методов выбора переменных для сокращения количества ковариатов. Методы выбора переменных могут эффективно определять наиболее влиятельные независимые характеристики интересующей переменной. Это помогает разработчику модели преодолеть проблему степеней свободы, которая является одной из наиболее часто встречающихся проблем при моделировании многомерных данных.

Большинство традиционных методов прогнозирования полагаются на подгонку данных к заранее заданным отношениям между зависимыми и независимыми переменными, таким образом предполагая определенный функциональный и стохастический процесс. Напротив, машинное обучение (МО) предлагает другой подход к статистическому анализу и прогнозированию, в частности, потому что он в значительной степени управляется данными, так как почти не делает предположений о базовых статистических отношениях в данных [1]. Основная цель МО – превратить информацию в знания и ценность, позволяя данным «говорить». Также МО – это термин, который обычно используют для обозначения процедур, которые в основном используются для точечного прогнозирования в условиях с неизвестной структурой.

Прогнозирование инфляции стало особенно актуальным после перехода России к режиму таргетирования инфляции в 2014 году. Это означает, что Банк России устанавливает конкретные цели по уровню инфляции и использует различные инструменты для их достижения. В современной экономической науке активно обсуждается роль методологий прогнозирования в контексте анализа

исторических данных для выявления закономерностей, предсказывающих будущие экономические тенденции.

Адекватное прогностическое моделирование выступает в качестве критического инструмента, способствующего предотвращению финансовых потрясений, совершенствованию стратегического управления, эффективной аллокации ресурсов и достижению выдающихся результатов в различных секторах экономики. Одним из передовых подходов в области прогностического анализа является модель «Случайный лес», которая зарекомендовала себя как один из лидирующих инструментов машинного обучения благодаря своей высокой предсказательной точности и универсальности в решении задач как классификации, так и регрессии. Модель превосходит как стандартные одномерные модели прогнозирования. Например, авторегрессионные модели (класс моделей МО, которые автоматически предсказывают следующий компонент последовательности на основе всех предыдущих компонентов этой последовательности и несколько других МО, включая регрессию лассо (LASSO, Least Absolute Shrinkage and Selection Operator) – это вариация линейной регрессии, специально адаптированная для данных, которые демонстрируют сильную мультиколлинеарность (высокую корреляцию признаков друг с другом), ридж-регрессию или гребневую регрессию (ridge regression – один из методов понижения размерности) [2].

Модель «Случайный лес» (random forest) находит свое применение в прогнозировании ключевых макроэкономических индикаторов, в том числе уровня инфляции, что подчеркивает ее значимость в экономической теории и практике. Принцип работы модели базируется на генерации ансамбля деревьев решений, каждое из которых обучается на отдельных подвыборках исходного набора данных, что обеспечивает разнообразие в обучающей выборке и уменьшает риск переобучения. Результаты работы отдельных деревьев интегрируются методом голосования или усреднения, что способствует повышению общей точности и стабильности прогнозов. Важно отметить, что модель позволяет проводить оценку значимости переменных, что представляет неоценимую пользу при анализе сложных данных и экономическом моделировании. Использование мо-

дели не ограничивается исключительно экономической сферой. Ее применение распространяется на медицину, биологические исследования и множество других областей, где требуется обработка больших объемов данных для решения задач классификации и регрессии, а также для аналитического исследования и оценки влияния различных факторов. Таким образом, модель «Случайный лес» демонстрирует свою многостороннюю применимость и эффективность, подтвержденную многочисленными исследованиями в области данных и машинного обучения

На практике одной из основных проблем с деревьями регрессии является их высокая дисперсия предсказаний. Обычно небольшое изменение в данных приводит к совершенно иной последовательности разделений. Основной причиной такой нестабильности является иерархическая природа алгоритма (эффект большой ошибки в верхнем разделении распространяется на все разделения ниже). Чтобы преодолеть эту проблему, можно использовать метод бэггинга (bootstrap aggregating, бутстреп-агрегирование), который заключается в подгонке одного и того же дерева несколько раз к версиям обучающих данных, отобранным методом бэггинга, а затем в усреднении результата. Такой подход часто улучшает производительность модели, поскольку он уменьшает дисперсию прогноза, не увеличивая слишком сильно смещение.

Фундаментальная концепция «Случайного леса» проста – это мудрость толпы (результат специфического процесса, в ходе которого независимые суждения статистически обобщаются (например, путем выявления среднего значения) для получения окончательного и наиболее точного решения или прогноза). Причина, по которой модель случайного леса работает так хорошо, заключается в том, что большое число относительно некоррелированных деревьев, работающих совместно, будет превосходить любую из их отдельных составляющих. Ключевым фактором является слабая корреляция между деревьями. Точно так же, как инвестиции с низкими корреляциями (например, акции и облигации) объединяются, чтобы сформировать портфель больший, чем сумма его частей, некоррелированные модели могут прогнозировать точнее, чем любой индивиду-

альный прогноз. Причина такого эффекта состоит в том, что деревья защищают друг друга от своих индивидуальных ошибок, по крайней мере до тех пор, пока они не будут постоянно ошибаться в одном и том же направлении. При этом некоторые деревья могут быть неправильными, но многие другие будут правильными, а значит, группа деревьев способна двигаться в верном направлении [3].

Итак, «Случайный лес» – это алгоритм классификации, состоящий из многих деревьев решений. Использование бэггинга при построении каждого отдельного дерева, помогает создать некоррелированный лес, прогноз которого точнее, чем у любого отдельного дерева.

Список литературы

1. Букина Т. Эффективность машинного обучения при прогнозировании инфляции / Т. Букина, Д. Кашин [Электронный ресурс]. – Режим доступа: <https://naked-science.ru/article/column/pri-prognozirovanii-infly> (дата обращения: 11.10.2024).

2. Иванова Л.Н. Методы машинного обучения для прогнозирования / Л.Н. Иванова, А.В. Куркин, С.Е. Иванов [Электронный ресурс]. – Режим доступа: <https://www.elibrary.ru/item.asp?id=44676935> (дата обращения: 11.10.2024).

3. Павлов Е. Прогнозирование инфляции в России с помощью нейронных сетей / Е. Павлов [Электронный ресурс]. – Режим доступа: <https://www.elibrary.ru/item.asp?id=42572917> (дата обращения: 11.10.2024).