Замещение системы Лоренца автономной ансамблевой моделью Extra Trees

Филимонова Елена Геннадиевна
кандидат экономических наук, доцент кафедры высшей математики и механики, Санкт-Петербургский государственный университет аэрокосмического приборостроения, elenaf3@yandex.ru,

Титов Вадим Евгеньевич
старший преподаватель кафедры высшей математики и механики, Санкт-Петербургский государственный университет аэрокосмического приборостроения, vadimtitov460@gmail.com,

Воронова Полина Андреевна
студентка, институт ФПТИ, Санкт-Петербургский государственный университет аэрокосмического приборостроения, poli93403@gmail.com

В работе рассматривается возможность замещения системы Лоренца автономной ансамблевой моделью машинного обучения на основе метода Extra Trees. Исходная динамическая система используется для формирования обучающих и тестовых данных, после чего модель функционирует независимо от исходных дифференциальных уравнений. Обучение проводится по зависимости между текущим состоянием системы и изменением ее фазовых координат за один временной шаг. После завершения обучения прогнозируемые состояния рекурсивно используются в качестве входных данных, что позволяет формировать автономную фазовую траекторию. Для оценки качества модели рассматриваются два режима: одношаговое прогнозирование и длительное автономное моделирование. Такой подход позволяет отдельно анализировать точность восстановления локального закона динамики и устойчивость модели при последовательном накоплении прогнозов. Особое внимание уделяется специфике хаотических систем, для которых высокое качество локального прогнозирования не гарантирует длительного совпадения конкретных траекторий. Оценка проводится на основе регрессионных метрик и сопоставления фазовых траекторий в трехмерном пространстве. Предлагаемый подход может использоваться для исследования возможностей ансамблевых методов при построении суррогатных моделей нелинейных динамических систем.

Ключевые слова: система Лоренца, хаотическая динамика, Extra Trees, ансамбль деревьев решений, машинное обучение, фазовая траектория, автономная модель, нелинейная динамическая система, прогнозирование динамики, фазовое пространство.

Введение

Исследование нелинейных динамических систем традиционно связано с построением математической модели процесса и последующим численным решением системы дифференциальных уравнений. Такой подход позволяет получать временные зависимости переменных и исследовать характер движения системы в фазовом пространстве. Вместе с тем аналитическое описание динамики доступно не для всех процессов. В экспериментальных задачах исследователь может располагать большим количеством наблюдений за состоянием объекта, не имея при этом достаточно точной математической модели его эволюции [2, 3].

Развитие методов машинного обучения позволяет рассматривать данные не только как источник информации для прогнозирования отдельных значений, но и как основу для построения альтернативной динамической модели. В такой постановке алгоритм обучается определять изменение состояния системы по ее текущим параметрам, после чего может использоваться автономно. Каждое рассчитанное состояние становится исходным для следующего шага, а обращение к первоначальной системе дифференциальных уравнений после завершения обучения не требуется.

Особый интерес представляет применение данного подхода к хаотическим системам. Их динамика характеризуется высокой чувствительностью к малым изменениям состояния, поэтому даже незначительная локальная погрешность модели способна со временем привести к существенному расхождению двух траекторий. Это делает недостаточным использование только традиционных метрик машинного обучения и требует отдельного анализа поведения модели при длительной автономной генерации последовательности состояний.

В качестве объекта исследования выбрана система Лоренца — один из классических примеров детерминированной хаотической динамики [1]. Она удобна для вычислительного эксперимента, поскольку позволяет получить эталонную фазовую траекторию и сопоставить ее с траекторией, сформированной моделью машинного обучения. В качестве метода аппроксимации используется Extra Trees — ансамблевый алгоритм, основанный на совокупности крайне рандомизированных деревьев решений, и позволяющий моделировать сложные нелинейные зависимости без предварительного задания их функционального вида.

В отличие от задачи обычного прогнозирования временного ряда в настоящем исследовании ансамбль Extra Trees используется как замена исходного механизма изменения состояния. Система Лоренца применяется для формирования обучающих и контрольных данных, после чего автономная фазовая траектория строится исключительно с помощью обученной ансамблевой модели [1, 4, 8].

Цель исследования состоит в оценке возможности использования Extra Trees для замещения динамики системы Лоренца и в установлении различий между точностью одношагового восстановления состояния и качеством длительного автономного воспроизведения фазовой траектории.

Построение автономной ансамблевой модели динамики системы Лоренца

Вычислительный эксперимент основывался на классической системе Лоренца

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = x(\rho - z) - y, \\ \dot{z} = xy - \beta z, \end{cases}$$

Где x, y и z – переменные состояния системы; σ, ρ и β – её параметры. В расчетах приняты значения σ=10, ρ=28, β=3/8 [1].

При данном сочетании параметров система демонстрирует хаотический режим и формирует характерный странный аттрактор в трехмерном фазовом пространстве.

Для получения исходных данных система Лоренца численно интегрировалась при нескольких начальных условиях. Использование нескольких траекторий было необходимо для более полного представления различных участков фазового пространства в обучающей выборке [6]. Итоговая выборка включала 30 000 состояний, каждое из которых описывалось тремя фазовыми координатами.

Текущее состояние системы представим вектором $X_k = (x_k, y_k, z_k)$.

Вместо непосредственного прогнозирования следующего состояния ансамблевая модель обучалась определять его изменение за один временной шаг:

$\Delta X_k = X_{k+1} - X_k$.

Таким образом, входными признаками являлись значения x_k, y_k, z_k.

Обучающая зависимость имела вид

$(x_k, y_k, z_k) \rightarrow (\Delta x_k, \Delta y_k, \Delta z_k)$.

Такой способ построения выборки позволяет алгоритму моделировать непосредственно локальное изменение состояния []7, 8. Для малого шага дискретизации соседние точки траектории находятся достаточно близко друг к другу, поэтому использование приращений делает задачу более ориентированной на восстановление направления и величины движения системы в фазовом пространстве.

Обученный ансамбль Extra Trees определяет оценку приращения

$$\hat{X}_{k+1} = \hat{X}_k + F_{ET}(\hat{X}_k).$$

Получается последовательность

$$\hat{X}_0 \rightarrow \hat{X}_1 \rightarrow \hat{X}_2 \rightarrow \ldots \rightarrow \hat{X}_N.$$

После начала этого процесса исходная система Лоренца для расчета модельной траектории больше не используется. Поэтому сформированное отображение можно рассматривать как самостоятельную дискретную динамическую систему.

По своей вычислительной структуре данный подход частично напоминает численную схему интегрирования. При традиционном решении дифференциальной системы изменение состояния определяется известной аналитической функцией. В рассматриваемом случае ее функциональная роль фактически передается ансамблю деревьев:

$$X_{k+1} = X_k + F_{ET}(X_k).$$

При этом модель не получает сведений об аналитическом виде исходных уравнений. Она не использует непосредственно выражения, содержащие произведения фазовых переменных, и не располагает информацией о физических или математических предпосылках системы Лоренца. Закон изменения состояния формируется только на основании обучающей совокупности наблюдений.

В этом состоит одно из существенных отличий предлагаемого подхода от методов идентификации параметров известной системы. Задача не заключается в определении неизвестных значений параметров при заранее заданной структуре уравнений. Ансамбль формирует альтернативный дискретный оператор перехода, не воспроизводя исходную аналитическую запись системы.

Вместе с тем автономное использование модели предъявляет значительно более жесткие требования, чем одношаговая регрессия. Ошибка определения очередного приращения приводит к появлению состояния, отличающегося от эталонного. Именно это измененное состояние становится входом следующего шага. Поэтому погрешности начинают последовательно влиять друг на друга.

Для обычного устойчивого процесса подобное накопление ошибок может происходить относительно медленно. В системе Лоренца ситуация осложняется хаотическим характером динамики: близкие состояния со временем способны формировать существенно различающиеся траектории. По этой причине точное восстановление отдельных переходов еще не означает, что ансамблевая модель будет длительно следовать той же конкретной траектории, что и исходная система.

В связи с этим результаты модели необходимо анализировать в двух взаимодополняющих режимах: как качество локального оператора перехода и как поведение автономной динамической системы.

Оценка точности модели и сопоставление фазовых траекторий

Результаты вычислительного эксперимента показали значительное различие между одношаговой точностью Extra Trees и результатами длительного автономного моделирования.

При одношаговой проверке модель каждый раз получала истинное текущее состояние системы Лоренца и должна была определить состояние на следующем временном шаге. Средняя абсолютная ошибка составила *MAE*=0,00452985, среднеквадратичная ошибка — *RMSE*=0,01233879, а коэффициент детерминации достиг R^2=0,99999785.

Столь высокое значение коэффициента детерминации и низкие абсолютные показатели ошибки свидетельствуют о том, что ансамбль Extra Trees с высокой точностью аппроксимировал локальное изменение состояния системы. При использовании истинной точки фазовой траектории модель практически точно определяет положение следующей точки.

Полученный результат имеет принципиальное значение. Он показывает, что ансамбль деревьев способен извлечь из набора наблюдений достаточно точную зависимость между фазовым состоянием и его изменением, несмотря на то что аналитический вид исходной системы непосредственно в модель не передавался.

Совершенно иные значения получены при автономном запуске, когда каждое спрогнозированное состояние становилось входом следующего шага.

Таблица 1
Показатели ошибки автономной модели Extra Trees

Показатель	x	y	z	Общее значение
MAE	9,182484	10,067447	10,689769	9,979900
RMSE	11,005835	12,218347	13,031287	12,113778
R^2	−0,908341	−0,850796	−1,493123	—

Общая средняя абсолютная ошибка автономной траектории составила 9,979900. Наименьшее значение *MAE* наблюдалось для координаты *x* — 9,182484, тогда как для *y* оно увеличилось до 10,067447, а для *z* — до 10,689769.

Аналогичная закономерность наблюдалась для среднеквадратичной ошибки. Для координаты *x* *RMSE* составила 11,005835, для *y* — 12,218347, для *z* — 13,031287. Общая среднеквадратичная ошибка достигла 12,113778. Следовательно, наиболее значительное поточечное расхождение в проведенном эксперименте наблюдалось по третьей фазовой координате.

Особенно показательными являются значения коэффициента детерминации. В автономном режиме они оказались отрицательными для каждой из трех фазовых переменных:

$$R_x^2 = -0,908341,$$
$$R_y^2 = -0,850796,$$
$$R_z^2 = -1,493123.$$

Отрицательные значения свидетельствуют о том, что после длительного автономного расчета модельная последовательность уже не воспроизводит синхронные значения конкретной исходной траектории. При стандартной трактовке регрессионной задачи подобные показатели означали бы низкое качество прогноза.

Однако применительно к системе Лоренца эти значения необходимо рассматривать совместно с одношаговыми результатами. Именно контраст между двумя режимами является наиболее существенным результатом исследования. Одношаговый коэффициент детерминации составил 0,99999785, то есть был практически равен единице, тогда как при автономной генерации значения для всех координат стали отрицательными.

Такое различие нельзя объяснить неспособностью ансамбля моделировать локальную динамику. Напротив, одношаговые метрики показывают, что локальный переход был восстановлен с высокой точностью. Причиной резкого снижения долгосрочной точности является рекурсивное накопление даже очень малых отклонений в сочетании с хаотическими свойствами системы Лоренца.

Если на некотором шаге модель формирует состояние $\bar{X}_k$ с небольшим отклонением от эталонного X_k, то уже следующий прогноз рассчитывается не из истинной, а из измененной точки фазового пространства. Далее отклонение снова влияет на следующий результат. Вследствие высокой чувствительности системы к малым изменениям две первоначально близкие траектории постепенно теряют синхронизацию.

Масштаб этого эффекта особенно хорошо показывает сопоставление значений *RMSE*. Одношаговая среднеквадратичная ошибка составляет 0,01233879, тогда как общая *RMSE* автономной траектории достигает 12,113778. Следовательно, долгосрочное поточечное расхождение оказывается приблизительно в 982 раза выше одношаговой ошибки. Разница почти в три порядка демонстрирует, насколько существенным оказывается накопление небольших локальных погрешностей при автономном моделировании хаотического процесса.

Дополнительно рассчитывалось евклидово расстояние между соответствующими точками исходной и модельной фазовых траекторий:

$$d_k = \sqrt{(x_k - \bar{x}_k)^2 + (y_k - \bar{y}_k)^2 + (z_k - \bar{z}_k)^2}.$$

Среднее расстояние составило

$$\bar{d} = 18,996281,$$

а максимальное значение достигло

$$d_{max} = 42,938870.$$

Полученные значения подтверждают потерю поточечной синхронизации между траекторией исходной системы и автономной траекторией Extra Trees.

При этом использование только расстояний между синхронными точками для хаотической системы является недостаточным. После расхождения двух реализаций одна траектория может находиться на одном крыле аттрактора, а другая — на другом. Расстояние между ними в этот момент будет значительным, хотя обе точки потенциально могут принадлежать области, характерной для одной и той же динамической системы.

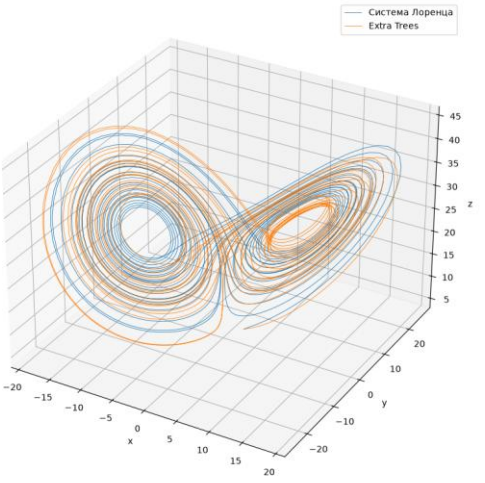


Рис.1. Сравнение фазовых траекторий системы Лоренца и автономной модели Extra Trees

Поэтому в работе наряду с численными показателями выполнено трехмерное сравнение фазовых траекторий системы Лоренца и модели Extra Trees.

На рисунке обе траектории представлены в общем пространстве координат *x*, *y* и *z*. Такое сопоставление позволяет дополнить традиционные показатели ошибки анализом геометрического поведения модели. Численные метрики отвечают на вопрос о том, находится ли Extra Trees в тот же момент времени в той же точке, что и исходная система. Фазовый портрет позволяет поставить другой вопрос — в какой степени модель после потери синхронизации продолжает воспроизводить характерную область движения системы Лоренца.

Это различие принципиально при оценке хаотических процессов. Для обычной задачи прогнозирования значительная *MAE* или отрицательный коэффициент детерминации являются достаточными основаниями говорить о

неудовлетворительном соответствии прогнозируемого временного ряда исходному. В рассматриваемой задаче такой вывод был бы неполным, поскольку целью является не только временная синхронизация, но и замещение оператора динамики.

Полученные результаты позволяют выделить два уровня качества ансамблевой модели.

Первый уровень связан с локальной точностью. По этому критерию Extra Trees продемонстрировал высокий результат. Одношаговая *MAE* составила 0,00452985, *RMSE* — 0,01233879, а коэффициент детерминации достиг 0,99999785. Следовательно, модель практически точно определяет ближайший переход между состояниями при условии, что получает корректную текущую точку.

Второй уровень связан с долгосрочной траекторной точностью. В данном случае модель не сохраняет синхронизацию с конкретной эталонной реализацией: общая *MAE* достигает 9,979900, *RMSE* — 12,113778, а коэффициенты детерминации оказываются отрицательными.

Таким образом, результаты не являются противоречивыми. Они характеризуют различные свойства модели. Extra Trees способен точно аппроксимировать локальный закон перехода, однако небольшие ошибки этого приближенного оператора при тысячах последовательных применений приводят к формированию иной реализации хаотического движения.

С научной точки зрения особенно важно, что результаты позволяют поставить под сомнение достаточность стандартного подхода к оценке машинных моделей динамических систем. Если бы качество Extra Trees оценивалось исключительно по одношаговому R^2), можно было бы сделать вывод о почти полном воспроизведении исходной системы. Если же использовать только долгосрочные значения R^2), вывод оказался бы противоположным. Ни один из этих вариантов отдельно не отражает полностью свойства модели.

Для оценки замещения хаотической системы необходимо одновременно учитывать локальную точность, характер накопления ошибки и геометрию фазовой траектории. Представленный в работе рисунок сравнения двух трехмерных траекторий поэтому следует рассматривать как часть экспериментального анализа, а не исключительно как иллюстрацию.

Дальнейшее развитие количественной оценки может быть связано с анализом геометрического сходства аттракторов независимо от синхронизации по времени. В частности, целесообразно сопоставлять распределения точек в фазовом пространстве, диапазоны и дисперсии переменных, расстояния между множествами точек, частоту переходов между крыльями аттрактора и показатели Ляпунова. Такие характеристики позволят определить, насколько автономная модель сохраняет именно динамический режим исходной системы после неизбежного расхождения отдельных траекторий.

Заключение

Проведенное исследование показало, что ансамбль Extra Trees способен с высокой точностью воспроизводить локальный закон изменения состояния системы Лоренца, однако при длительном автономном моделировании накопление малых ошибок приводит к расхождению с конкретной исходной траекторией.

Для хаотических систем это означает необходимость разделять качество локального восстановления динамики и качество долгосрочного поточечного прогноза. Оценка подобных моделей должна учитывать не только стандартные регрессионные показатели, но и устойчивость автономного режима, характер накопления ошибки и сохранение структуры движения в фазовом пространстве.

Перспективным направлением дальнейшего исследования является количественная оценка геометрического сходства исходного и восстановленного аттракторов, а также сравнение Extra Trees с другими ансамблевыми и нейросетевыми методами.

Литература

1. Lorenz E. N. Deterministic Nonperiodic Flow // Journal of the Atmospheric Sciences. 1963. Vol. 20, no. 2. P. 130–141. DOI: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.
2. Анищенко В. С. Сложные колебания в простых системах: механизмы возникновения, структура и свойства динамического хаоса в радиофизических системах. 2-е изд., доп. М.: Либроком, 2009. 314 с.
3. Breiman L. Random Forests // Machine Learning. 2001. Vol. 45. P. 5–32. DOI: 10.1023/A:1010933404324.
4. Geurts P., Ernst D., Wehenkel L. Extremely Randomized Trees // Machine Learning. 2006. Vol. 63. P. 3–42. DOI: 10.1007/s10994-006-6226-1.
5. Brunton S. L., Proctor J. L., Kutz J. N. Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems // Proceedings of the National Academy of Sciences. 2016. Vol. 113, no. 15. P. 3932–3937. DOI: 10.1073/pnas.1517384113.
6. Pathak J., Hunt B., Girvan M., Lu Z., Ott E. Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach // Physical Review Letters. 2018. Vol. 120, no. 2. Article 024102. DOI: 10.1103/PhysRevLett.120.024102.
7. Vlachas P. R., Pathak J., Hunt B. R., Sapsis T. P., Girvan M., Ott E., Koumoutsakos P. Backpropagation Algorithms and Reservoir Computing in Recurrent Neural Networks for the Forecasting of Complex Spatiotemporal Dynamics // Neural Networks. 2020. Vol. 126. P. 191–217. DOI: 10.1016/j.neunet.2020.02.016.
8. Смирнов А. О., Дик О. Е., Фролова Е. А., Титов В. Е. Исследование микроэлектромеханического акселерометра как нелинейной динамической системы // Наукоемкие технологии. 2021. Т. 22, № 8. С. 80–86. DOI: 10.18127/j19998465-202108-12.

Surrogate Modeling of the Lorenz System Using an Autonomous Extra Trees Ensemble

Filimonova E.G., Titov V.E., Voronova P.A.
Saint Petersburg State University of Aerospace Instrumentation

JEL classification: C10, C50, C60, C61, C80, C87, C90

The paper considers the possibility of replacing the Lorenz system with an autonomous ensemble machine learning model based on the Extra Trees method. The original dynamical system is used to generate training and test data, after which the model operates independently of the original differential equations. The model is trained using the relationship between the current system state and the change in its phase variables over one time step. After training, the predicted states are recursively used as new inputs, allowing an autonomous phase trajectory to be generated. The model is evaluated in two modes: one-step prediction and long-term autonomous simulation. This approach makes it possible to separately analyze the accuracy of local dynamics reconstruction and the stability of the model under recursive prediction. Particular attention is paid to the specific properties of chaotic systems, for which high local prediction accuracy does not necessarily ensure long-term agreement between individual trajectories. The evaluation is based on regression metrics and three-dimensional comparison of phase trajectories. The proposed approach can be used to investigate the applicability of ensemble methods to the construction of surrogate models for nonlinear dynamical systems.

Keywords: Lorenz system, chaotic dynamics, Extra Trees, decision tree ensemble, machine learning, phase trajectory, autonomous model, nonlinear dynamical system, dynamics prediction, phase space.

References

1. Lorenz E. N. Deterministic Nonperiodic Flow // Journal of the Atmospheric Sciences. 1963. Vol. 20, No. 2. P. 130–141. DOI: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.
2. Anishchenko V. S. Complex Oscillations in Simple Systems: Mechanisms of Emergence, Structure and Properties of Dynamical Chaos in Radiophysical Systems. 2nd ed. Moscow: Librokom, 2009. 314 p.
3. Breiman L. Random Forests // Machine Learning. 2001. Vol. 45. P. 5–32. DOI: 10.1023/A:1010933404324.
4. Geurts P., Ernst D., Wehenkel L. Extremely Randomized Trees // Machine Learning. 2006. Vol. 63. P. 3–42. DOI: 10.1007/s10994-006-6226-1.
5. Brunton S. L., Proctor J. L., Kutz J. N. Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems // Proceedings of the National Academy of Sciences. 2016. Vol. 113, No. 15. P. 3932–3937. DOI: 10.1073/pnas.1517384113.
6. Pathak J., Hunt B., Girvan M., Lu Z., Ott E. Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach // Physical Review Letters. 2018. Vol. 120, No. 2. Article 024102. DOI: 10.1103/PhysRevLett.120.024102.