

Методы оценивания качества моделей и полимодельных комплексов

Д. Е. Бунькова

Финансовый университет при Правительстве Российской Федерации (Финуниверситет), Financial University
debunkova@yandex.ru

Аннотация. В современных условиях назрела острая необходимость разработки методологических и методических основ теории оценивания и управления качеством моделей или, по-другому, квалиметрии моделей. Данная теория, являющаяся составной частью такой научной отрасли знаний как качествоведение, сама может быть декомпозирована на множество частных прикладных теорий, в которых должно проводиться оценивание качества моделей, используемых в определённой предметной области. Построение и последующая проверка работоспособности полученных моделей представляет собой сложный и итеративный процесс, по итогам которого достигается приемлемый уровень уверенности исследователя в том, что результаты, получаемые с помощью итоговой модели, окажутся практически полезными.

Ключевые слова: модели и моделирование, квалиметрия моделей, полимодельные комплексы, принятие решений, оценка качества

Анализ и оценивание моделей и их свойств происходит в процессе моделирования, которое выступает в качестве одного из целенаправленных процессов. В связи с этим возможно выделить такие исследования, в рамках которых производится рассмотрение проблем анализа и оценивания качества технологий моделирования и проблемы определения оптимальных технологий создания (моделирования) как отдельных моделей, так и полимодельных комплексов. Важное значение имеет также и эффективность технологии системного моделирования, которая может быть подразделена на такие составляющие, как: оперативность, результативность и ресурсоемкость. Оперативность представляет собой затраты временного характера для выполнения моделирования в целях получения установленного целевого эффекта. Результативность определяется объемом и качеством информации, получаемой в процессе моделирования, а также способностью моделирования достигать конкретной цели (конечного эффекта). Ресурсоемкость характеризует затраты всех видов возможных ресурсов при применении указанной технологии моделирования.

Развитие квалиметрии осуществляется по двум базовым направлениям: изучение общих вопросов квалиметрии моделей и управление качеством модели (полимодельного комплекса).

На протяжении последних десятилетий проблемы, которые связаны с оцениванием качеством продукции, стали предметом интенсивных исследований, проводимых

в такой новой научной отрасли знаний, как качествоведение, в которой изучаются закономерности получения и обработки информации о качестве объекта на всех этапах его жизненного цикла.

При этом выделяются следующие стандартные шаги, которые необходимо подробно рассмотреть.

1 шаг. Разведочный анализ данных (exploratory data analysis), главная цель которого – изучение статистических свойств имеющихся в наличии выборок (распределение переменных, наличие выбросов, необходимость трансформации и др.) и выявление характера взаимосвязей между откликом и предикторами.

2 шаг. Выбор методов построения моделей и спецификация систематической части последних. Например, модели типа «доза-эффект» в биологии на одном и том же исходном материале могут быть построены с использованием самых различных функций: логистической, экспоненциальной, Вейбулла, Гомперца, Михаелиса-Ментен, Брейна-Кузенса и т.д.

3 шаг. Оценка параметров моделей и их диагностика (Мастицкий, Шитиков, 2014). Диагностика и оценка валидности (model validation) включает в себя ряд стандартных процедур. Так, в случае с классическими регрессионными моделями, подгоняемыми по методу наименьших квадратов, выполняются: а) проверка статистической значимости модели в целом и анализ неопределенности оцененных коэффициентов; б) проверка допущений в отношении остатков модели; в) обнаружение необычных наблюдений и выбросов; г) построение графиков, позволяющих оценить соответствие модели структуре анализируемых данных.

4 шаг. Анализ вклада отдельных предикторов и селекция оптимальной комбинации из них (model selection). Оценка качества каждой модели-претендента (model evaluation) по совокупности объективных критериев эффективности, включая тестирование на порции “свежих” данных, не участвовавших в процессе оценивания коэффициентов.

5 шаг. Ранжирование нескольких альтернативных моделей и, при необходимости, подстройка их важнейших параметров (model tuning). Рассматриваемые в этом разделе параметрические регрессионные модели в классическом представлении являются аппроксимацией математического ожидания отклика Y по обучающей

выборке с помощью неизвестной функции регрессии $f(\dots)f(\dots)$:

$$E(Y|x_1, x_2, \dots, x_m) = f(\beta, x_1, x_2, \dots, x_m) + \epsilon,$$

$$E(Y|x_1, x_2, \dots, x_m) = f(\beta, x_1, x_2, \dots, x_m) + \epsilon,$$

где остатки ϵ отражают ошибку модели, т.е. необъяснимую случайную вариацию наблюдаемых значений зависимой переменной относительно ожидаемого среднего значения.

С практической точки зрения, тестирование таких моделей ставит своей задачей выявление следующих основных проблем их возможного использования:

Смещение (bias), или систематическая ошибка модели (сдвиг предсказываемых значений на некоторую трудно объяснимую величину);

Высокая случайная дисперсия прогноза, определяемая чаще всего излишней чувствительностью модели к небольшим изменениям в распределении обучающих данных;

Неадекватность – тенденция модели не отражать основных закономерностей генеральной совокупности данных и основываться на случайных флуктуациях обучающей выборки;

Переусложнение модели (overfitting), которое «так же вредно, как и ее недоусложнение».

Действительно, для любого проверочного наблюдения x_0 математическое ожидание среднеквадратичной ошибки его прогноза можно разложить на сумму трех величин: дисперсии $f(x_0)$, квадрата смещения $f(x_0)$ и дисперсии остатков ϵ (подробнее:

$$E[y_0 - f(x_0)]^2 = Var[f(x_0)] + [Bias(f(x_0))]^2 + Var(\epsilon),$$

$$E[y_0 - f(x_0)]^2 = Var[f(x_0)] + [Bias(f(x_0))]^2 + Var(\epsilon),$$

где Bias означает смещение, а Var – дисперсию. Здесь мы предположили, что неизвестная истинная функция $f(\dots)$ была оценена на большом числе обучающих выборок, а отклонения y_0 вычислялись по каждой из множества моделей с последующим усреднением результатов.

Из приведенного уравнения следует, что для минимизации ожидаемой ошибки прогноза необходимо подобрать такую модель, для которой одновременно достигаются низкое смещение и низкая дисперсия. Обратите внимание, что дисперсия никогда не может быть ниже некоторого уровня неустранимой ошибки $Var(\epsilon)$.

Существует также и феномен переусложнения модели, при котором наблюдается низкая дисперсия прогноза на обучающей совокупности, но часто получаются непредсказуемые результаты при тестировании блоков «свежих» данных, не участвовавших в построении модели.

I. СВОЙСТВА МОДЕЛЕЙ И ПОЛИМОДЕЛЬНЫХ КОМПЛЕКСОВ, ПОДЛЕЖАЩИЕ ОЦЕНКЕ

В современном анализе данных существует огромное количество свойств моделей, которые обязаны быть

проанализированы в рамках квалиметрии моделей таковыми являются надёжность, унификация, простота, открытость и доступность модели, их интеллектуальность, эффективность машинной реализации, сложность, идентифицируемость, устойчивость, чувствительность, управляемость, наблюдаемость моделей, их инвариантность, развиваемость и многие другие. Но стоит выделить ряд основных свойств, оказывающих существенное влияние на оценку моделей.

а) Адекватность. Указанным свойством должна обладать модель по отношению к тем или иным аспектам объекта-оригинала. Конечно, сталкиваясь с этим свойством на практике, предполагается лишь адекватность в некоторой степени, а не полная адекватность. Если брать сложные системы (рис. 1), то для одной модели характерно отражение лишь какой-либо стороны, аспекты прототипа, следовательно, понятия адекватности для такой модели не будет возникать, можно говорить лишь об адекватности указанного аспекта.



Рис. 1. Технология моделирования сложных систем

Для полимодельного комплекса может ставиться задача достижения адекватности в более широком смысле, охватывающем различные стороны прототипа. Однако во всех случаях оценивание адекватности модели (полимодельного комплекса) должна проводиться с учётом того, в какой степени на данной модели могут быть достигнуты цели моделирования (цели субъекта).

б) Простота и оптимальность модели (полимодельного комплекса). Для свойства адекватности модели характерна некая связь с такими свойствами, как простота и оптимальность модели. В самом деле, для того, чтобы достичь требуемой степени адекватности, иногда приходится идти на существенное усложнение модели. С другой стороны, если существует возможность выбора между различными моделями, имеющими примерно одинаковую адекватность, целесообразно из них выбрать наиболее простую. Перечисленные вопросы приобретают особую актуальность при оптимальном выборе структуры полимодельного комплекса. В данном случае адекватность моделирования уже определяется не только свойствами каждой модели в отдельности, но также характеристиками взаимодействия моделей.

в) Гибкость (адаптивность) моделей. Данное свойство моделей предполагает ввод в состав моделей таких параметров и структур, которые можно менять в заданных диапазонах для достижения целей моделирования.

з) *Универсальность и проблемная ориентация моделей.* Многочисленные исследования, направленные на поиск указанного компромисса, показали, что в настоящее время разработка универсальных моделей, ориентированных на широкую предметную область, является трудно разрешимой проблемой. Целесообразно создавать модели, специализированные по допустимому классу моделируемых объектов и универсальные по поддерживаемым функциям.

II. ГРАФИЧЕСКИЙ АНАЛИЗ КАК ПРИМЕР МЕТОДА ОЦЕНИВАНИЯ МОДЕЛЕЙ И ПОЛИМОДЕЛЬНЫХ КОМПЛЕКСОВ

Самый простой метод оценки полученной модели – линейный график. На него достаточно нанести фактические значения, расчётные и прогнозные. Так мы сможем увидеть, как модель описала данные и насколько полученный прогноз соответствует сложившейся динамике. Более того, так можно понять (на основе знаний о прогнозируемом процессе), насколько полученные прогноз правдоподобен.

Очень простой пример с линейным графиком показан рис. 2.

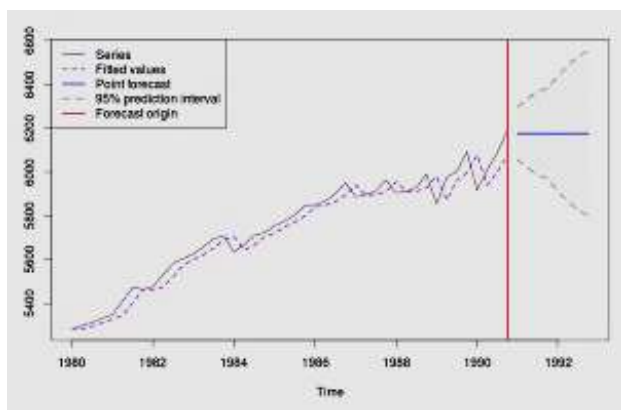


Рис. 2. Линейный график по несезонной модели ETS (M,N,N)

На нём показан ряд фактических значений (чёрная сплошная линия «Series»), ряд расчётных (фиолетовая пунктирная линия «Fitted values»), точечный прогноз (синяя сплошная линия «Point forecast») и 95% доверительный интервал («95% prediction interval»). Уже глядя на этот график, можно сделать несколько выводов:

- Ряд фактических значений был описан моделью неплохо, хотя основные характеристики ряда выявлены не были – это видно, потому что расчётные значения всё время как будто отстают от фактических на один шаг. К тому же ряд расчётных значений не такой гладкий, как хотелось бы и в ряде наметились элементы сезонности (это особенно чётко видно с 1986 года) – каждый четвёртый квартал виден пик показателя. Это не криминал, но может говорить о том, что стоило бы обратиться к другой модели либо другому методу оценивания.

- Прогноз по нашей модели представляет собой прямую линию, параллельную оси абсцисс. Это хороший прогноз в случае, если ряд данных не имеет явной тенденции к росту либо снижению. Однако в нашем случае ряд данных явно демонстрирует увеличение показателя во времени. Поэтому такая тенденция в будущем хоть и возможна, но, скорее всего, не соответствует ожиданиям. Для того, чтобы здесь сделать какой-то однозначный вывод, нужно, конечно, понимать, с каким показателем мы имеем дело, и насколько реалистичен такой сценарий на практике.
- Прогнозные интервалы значительно расширяются и к последнему наблюдению составляют 5800 – 6500. Опять же, такие широкие интервалы – это не страшно, но одна из потенциальных причин этой ширины – неправильно выбранная (или специфицированная) модель. Так же на возможно неправильно выбранную модель указывает неровность прогнозного интервала при таком ровном прогнозе. Возможно, всё-таки надо было использовать другую модель (с трендом и сезонностью), а, может быть, мы не учли какие-то важные факторы.

На основе этих пунктов уже можно заключить, что стоит попробовать ещё какую-нибудь модель для прогнозирования этого ряда данных. Если мы используем для выбора модели процедуру ретропрогноза, то на тот же самый график стоит нанести значения из проверочной выборки. Другая модель, построенная по тому же ряду данных, показана на рис. 3.

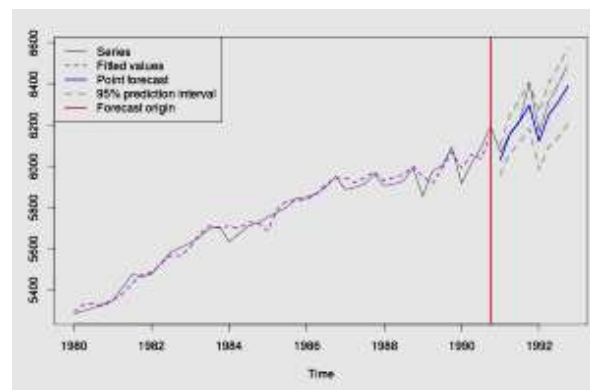


Рис. 3. Линейный график по сезонной модели ETS (M,A,M)

В общем, всё в этом графике указывает на то, что стоит обратиться к модели с трендом и сезонностью. В ряде случаев по одному лишь линейному графику бывает сложно сделать выводы о качестве модели. Например, те же самые выбросы оценить по нему на глаз непросто. В этом случае нужно обратиться к графикам по остаткам модели. Даже простой линейный график по остаткам может дать много полезной информации.

Однако прежде, чем приступать к непосредственному анализу, надо понять, что нам нужно. В идеальной модели остатки ни от чего не зависят, выглядят распределёнными случайно и желателно даже нормально. Графически это

должно выражаться отсутствием какой-либо предсказуемости в остатках и ровными гистограммами и ящичковыми диаграммами.

На следующих рисунках приведены три графика по остаткам модели ETS(M,N,N) (той, которая соответствует первому графику в этой статье).

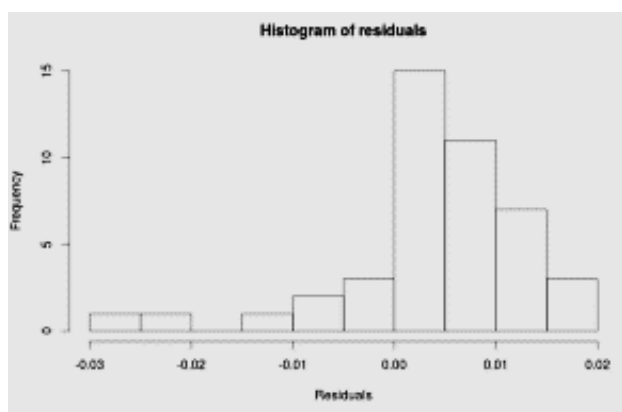


Рис. 4. Гистограмма по остаткам модели ETS(M,N,N)

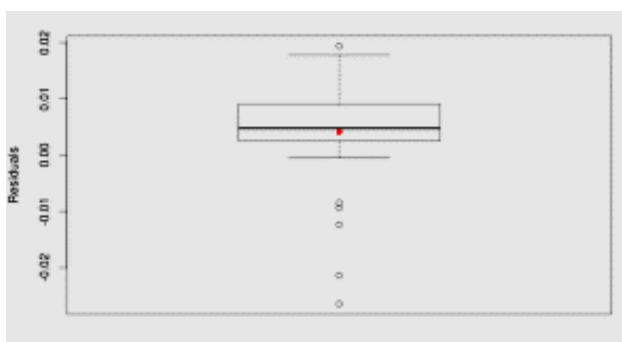


Рис. 5. Ящичковая диаграмма по остаткам модели ETS(M,N,N)

Взглянем критически на эти графики и попробуем вынести какую-нибудь полезную информацию.

- Линейный график по остаткам демонстрирует периодические колебания (которые особенно заметны примерно с 1986 года) – это всё из-за того, что мы не учли в модели сезонность. По остаткам она видна более явно, чем по графику по исходному ряду.
- Гистограмма по остаткам демонстрирует явную асимметрию: очень много ошибок лежит выше нуля (справа). Это так же указывает на неправильную спецификацию модели. Обычно исследователи при построении разных моделей стараются добиваться нормально распределённых остатков, но это требование на практике очень сложно удовлетворить. Именно поэтому нам нужно, чтобы они были хотя бы симметрично распределены относительно нуля. Это будет указывать на то, что модель построена без систематических ошибок.

- Ящичковая диаграмма показывает то же, что и гистограмма: мы имеем дело с асимметричным распределением. Помимо этого, диаграмма показала, что практически все отрицательные остатки можно статистически считать выбросами – то есть они настолько редко случаются, что играют небольшую роль в описательной способности модели. На себя так же обращает внимание завышенное среднее значение ошибки (красная точка по середине ящичка) – это так же сигнализирует о неправильной спецификации.

Как видим, на основе такого небольшого анализа можно получить много полезной информации об аппроксимационных свойствах модели и принять решение о том, что делать с моделью либо к какой модели обратиться. Ошибка прогноза измеряется чаще всего в значениях MAPE и MAE. Данный подход позволяет сравнить собственные модели, свои результаты с чужими. В научных статьях чаще всего используются именно эти показатели.

Время вычисления прогноза является показательным и даже критичным, когда прогноз нужно делать быстро. Например, прогнозировать вперед на 5 минут каждые 5 минут. В таком случае, если модель работает 6 минут, ее нельзя применять, то есть она неэффективна. В остальных случаях этот показатель больше отражает удобство использования модели. В задачах машинного обучения для оценки качества моделей и сравнения различных алгоритмов используются метрики, а их выбор и анализ – неперемная часть длительной работы.

В современных условиях человек научился рационально подходить и оценивать все выполняемые процессы. При этом особое значение в настоящее время уделяется таким процессам, как планирование и управление процессами. Поэтому в науке и практической деятельности появилась специфическая новая сфера деятельности, предметом которой является решение большого количества организационных и управленческих задач, – системный анализ, составной частью которого выступает системное моделирование во всем его многообразии.

СПИСОК ЛИТЕРАТУРЫ

- [1] Беляева М.А. Системный анализ технологий и бизнес-процессов в производстве: монография / М.А. Беляева. М.: РЭУ им. Г.В. Плеханова, 2015. с. 321-353
- [2] Вдовин В.М. Теория систем и системный анализ: Учеб. для бакалавров / В.М. Вдовин, Л.Е. Суркова и др. М.: Дашков и К, 2016. 644 с.
- [3] Звягин Л.С. Метасистемный подход в управлении как фундаментальный фактор формирования процессов хозяйственного развития // Хроноэкономика. 2018. № 2 (10). С. 11-17.
- [4] Звягин Л.С. Мягкие имитационные модели региональных экономических систем // Мягкие измерения и вычисления. 2018. № 10. С. 41-46.
- [5] Звягин Л.С. Эффективность функционирования систем и анализ критериев безопасности // Международная конференция по мягким вычислениям и измерениям. 2018. Т.1. С. 255-260.
- [6] Коршунов Г.И., Фрейман В.И. Модели и методы оценки соответствия показателей качества продукции и результативности подготовки специалистов // Фундаментальные исследования. 2015. № 12-6. С. 1116-1120; URL: <https://www.fundamental-research.ru/ru/article/view?id=39740> (дата обращения: 13.03.2019).