

МОДЕЛИ ПРОГНОЗИРОВАНИЯ НА ОСНОВЕ РАСПРЕДЕЛЕНИЯ ПУАССОНА В РЕАЛЬНОМ ВРЕМЕНИ

К.Р. Сварян, Н.Ю. Юдина

ФГБОУ ВО «Воронежский государственный лесотехнический университет
имени Г.Ф. Морозова»

Аннотация. В статье будут исследованы основные теоретические аспекты распределения Пуассона и его применение для прогнозирования событий. Рассмотрим процесс разработки таких моделей, алгоритмы для обработки данных в реальном времени и программные инструменты, обеспечивающие их реализацию. Кроме того, обсудим примеры использования этих моделей в реальных условиях, а также их преимущества и ограничения.

Ключевые слова: распределение Пуассона, прогнозирование, обработка данных, алгоритмы, потоковые данные, модели, статистический анализ.

REAL-TIME FORECASTING MODELS BASED ON THE POISSON DISTRIBUTION

K.R. Svarian, N.Yu. Yudina

Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. The article will explore the main theoretical aspects of the Poisson distribution and its application for event forecasting. We will examine the process of developing such models, algorithms for real-time data processing, and software tools that enable their implementation. Additionally, we will discuss examples of using these models in real-world conditions, as well as their advantages and limitations.

Keywords: Poisson distribution, forecasting, data processing, algorithms, streaming data, models, statistical analysis

В условиях стремительного роста объема данных и увеличения потребности в оперативном анализе, прогнозирование событий в реальном времени стало важной задачей для различных отраслей. Компании, занимающиеся управлением трафиком, медицинские учреждения, службы поддержки и многие другие организации нуждаются в инструментах, позволяющих эффективно предсказывать вероятностные события. Одним из подходов, доказавших свою эффективность в данной области, является использование распределения Пуассона для моделирования и прогнозирования.

Распределение Пуассона, описывающее количество событий, происходящих за фиксированный интервал времени, широко применяется в ситуациях, где события происходят случайно и независимо друг от друга. Эта

модель находит применение в таких задачах, как прогнозирование числа обращений в службу поддержки, оценка частоты отказов оборудования или предсказание посещаемости веб-сайтов. Благодаря своей простоте и способности точно описывать вероятностные события, распределение Пуассона стало незаменимым инструментом для многих аналитиков и исследователей.

Цель данной статьи – изучить возможности применения моделей на основе распределения Пуассона для прогнозирования событий в реальном времени. Мы рассмотрим теоретическую основу данной модели, проанализируем алгоритмы, позволяющие реализовать прогнозирование в условиях потоковой обработки данных, а также исследуем реальные примеры использования этой модели в различных сферах. Кроме того, в статье будут обсуждены преимущества и ограничения подхода, что позволит оценить его потенциал для решения различных прикладных задач.

Распределение Пуассона описывает вероятность получения k успехов за заданный интервал времени.

Итак, если количество испытаний n достаточно велико, а вероятность появления события в отдельно взятом испытании весьма мала то вероятность того, что в данной серии испытаний событие появится ровно k раз, можно приближенно вычислить по формуле Пуассона

$$P_n(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \quad (1)$$

где λ – среднее количество успехов за определенный интервал;

k – количество успехов;

e – константа, равная приблизительно 2,71828.

Несмотря на широкое применение, распределение Пуассона имеет определенные ограничения. В реальных условиях эти предпосылки могут быть нарушены, например, в случае зависимости между событиями или изменяющейся во времени интенсивности. Такие ситуации требуют модификации модели или использования других распределений (например, распределения сдвинутого Пуассона) для более точного прогнозирования.

Одним из ключевых применений распределения Пуассона является прогнозирование количества обращений в службы поддержки клиентов. Такие модели помогают предсказать, сколько звонков или заявок может поступить в определенный период времени, что позволяет оптимизировать распределение ресурсов, таких как операторы колл-центра или технические специалисты.



Рисунок 1 – Вероятное количество обращений при среднем количестве 5

Рестораны используют распределение Пуассона для моделирования количества ожидаемых клиентов, которые будут приходить в ресторан в день.

Например, предположим, что данный ресторан принимает в среднем 100 клиентов в день (но точно не больше 150). Мы можем использовать распределение Пуассона, чтобы найти вероятность того, что ресторан примет больше, чем определенное количество клиентов.

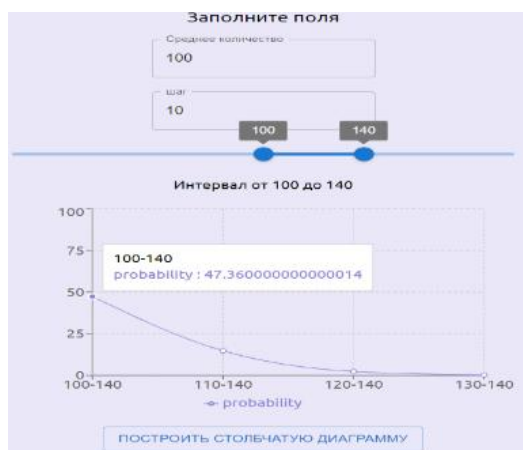


Рисунок 2 – Вероятное количество посетителей ресторана

Для оценки эффективности моделей прогнозирования на основе распределения Пуассона могут быть использованы различные метрики, такие как:

Среднеквадратическая ошибка (MSE) – измеряет среднее квадратическое отклонение прогнозируемых значений от фактических. Этот показатель чувствителен к большим ошибкам, поэтому он особенно полезен, когда важно минимизировать значительные отклонения в прогнозе.

Средняя абсолютная ошибка (MAE) – рассчитывает среднее абсолютное отклонение прогнозов от фактических значений. MAE предоставляет более интуитивно понятную меру ошибки, так как измеряется в тех же единицах, что и сами прогнозы.

Ошибка процентного отклонения (MAPE) – оценивает средний процентный отклонение прогнозов от фактических значений. Эта метрика удобна для сравнений между различными моделями и разными наборами данных, поскольку нормализует ошибки.

Особую сложность в оценке таких моделей представляет их работа в реальном времени, где необходимо учитывать динамическую изменчивость данных. Например, интенсивность события может варьироваться в зависимости от времени суток, дня недели или других временных факторов. Поэтому, чтобы правильно оценить модель в реальных условиях, следует рассматривать:

Временные окна прогнозирования: Прогнозирование на коротких временных интервалах (например, минутные или часовые данные) требует от модели способности быстро адаптироваться к изменениям.

Обновление модели: Регулярное обновление параметров модели на основе поступающих данных помогает сохранить её актуальность и точность. Оценка эффективности моделей на основе распределения Пуассона в реальном времени требует комплексного подхода, включающего использование различных метрик, адаптацию к динамическим изменениям и сравнение с альтернативными моделями. Успешное применение таких моделей позволяет существенно улучшить точность прогнозов и повысить эффективность решений, принимаемых на их основе.

Список литературы

1. Крицкий О.Л. Теория вероятностей и математическая статистика для технических университетов. I. Теория вероятностей: учебное пособие / О.Л. Крицкий, А.А. Михальчук, А.Ю. Трифонов, М.Л. Шинкеев; Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2010. – 212 с
2. Зорич В. А. Математический анализ. Часть I. – Изд. 10-е, испр. – М.: МЦНМО, 2019. – xii+564 с. ISBN 978-5-4439-4029-8, 978-5-4439-4030-4 (ч. I).
3. Печинкин А.В., Тескин О.И., Цветкова Г.М. и др. Теория вероятностей : учеб. для вузов. – 3-е изд., исправ. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2004. – 456 с.
4. Krishna Kumar, B. An M/G/1/1 queue with unreliable server and no waiting capacity / B. Krishna Kumar, D. Arivudainambi, A. Vijayakumar // Inf. Manage. Sci. – 2022. – Vol. 13. – P. 35–50.
5. Optimization analysis of an unreliable multi-server queue with a controllable repair policy / Wu C.-H., Lee W.-C., Ke J.-C., Liu T.-H. // Computers and Operations Research. – 2021. – № 49. – pp. 83-96.

References

1. Kritsky O.L. et al. Probability Theory and Mathematical Statistics for Technical Universities. I. Probability Theory: A Textbook / O.L. Kritsky, A.A. Mikhailyuk, A.Yu. Trifonov, M.L. Shinkeev; Tomsk Polytechnic University. – Tomsk: Tomsk Polytechnic University Publishing, 2010. – 212 p.
2. Zorich V.A. Mathematical Analysis. Part I. – 10th ed., revised. – Moscow: MCCME, 2019. – xii+564 p. ISBN 978-5-4439-4029-8, 978-5-4439-4030-4 (Part I).
3. Pechinkin A.V., Teskin O.I., Tsvetkova G.M. et al. Probability Theory: A Textbook for Universities – 3rd ed., revised. – Moscow, 2004. – 456 p.
4. Krishna Kumar, B. An M/G/1/1 queue with unreliable server and no waiting capacity / B. Krishna Kumar, D. Arivudainambi, A. Vijayakumar // Inf. Manage. Sci. – 2022. – Vol. 13. – P. 35–50.
5. Optimization analysis of an unreliable multi-server queue with a controllable repair policy / Wu C.-H., Lee W.-C., Ke J.-C., Liu T.-H. // Computers and Operations Research. – 2021. – № 49. – P. 83-96.