БАЙЕСОВСКОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ НА ОСНОВЕ ЧИСЛЕННЫХ МЕТОДОВ МОНТЕ-КАРЛО

Лобач В. И.

Белорусский государственный университет ФПМИ, Минск, Беларусь,
e-mail: lobach@bsu.by

Машинное обучение представляет собой раздел искусственного интеллекта, который базируется на идее того, что системы могут обучаться на основе поступающих данных и знаний, выявлять закономерности, строить прогнозы и принимать решения с минимальным вмешательством человека. В последние годы все большее влияние на машинное обучение оказывает байесовский подход [1,2]. Он позволяет строить адекватные модели в условиях малых выборок, а также учитывать экспертные знания ещё до получения данных. Байесовский подход базируется на принципе объединения априорных предположений о распределении параметров рассматриваемой модели и эмпирических данных для получения апостериорного распределения вероятностей параметров рассматриваемой модели и обладает рядом преимуществ перед классическими методами. Этот подход предоставляет возможность автоматической настройки структурных параметров алгоритмов машинного обучения, обеспечивает корректную работу с данные, достоверность которых установлена не точно.

Рассмотрим вероятностное распределение p(X) (плотность распределения вероятностей одномерной или многомерной случайной величины). Метод Монте–Карло предполагает генерацию выборки из этого распределения

$$X_1, X_2, ..., X_N \quad \sim \; p(X).$$

Данная выборка может быть использована для оценки вероятностных интегралов вида

$$E\{f(X)\} = \int_{-\infty}^{\infty} f(X)\, p(X)\, dX \approx \frac{1}{N}\sum_{i=1}^{N} f(X_i), \tag{1}$$

где f(X) – некоторая известная функция от X. К интегралам такого вида сводятся многие шаги при осуществлении байесовского подхода в исследовании вероятностных моделей. Например, такими интегралами являются функционал E(ln(p(X,T/θ))) на M – шаге EM – алгоритма, байесовская оценка неизвестного
параметра θ вероятностного распределения

$$E(\theta / X) = \int_{-\infty}^{\infty} \theta\, p(\theta / X)\, d\theta$$

(условное математическое ожидание, которое является оптимальной в среднеквадратическом смысле оценкой параметра θ) и т.д.

Байесовский подход в статистическом анализе основывается на теореме Байеса. Пусть θ – вектор параметров модели, а X – имеющийся набор данных. Через p(θ) обозначим априорное распределение параметра θ (априорная информация о параметрах, то есть информация, полученная до проведения эксперимента). В процессе наблюдений эта информация постепенно уточняется. Ставится задача оценивания неизвестного значения величины θ, наблюдая ее косвенные характеристики, после чего для решения проблемы прогнозирования применяются прогнозирующие статистики на основе подстановочного принципа (вместо неизвестного параметра θ используются его оценки). Априорное распределение вероятностей пересматривается на основе существующей реализации временного ряда X для получения апостериорной плотности распределения вероятностей p(θ/X). По формуле Байеса устанавливается связь между p(θ/X), p(X/θ), p(θ)

$$p(\theta / X)=\frac{p(X / \theta) p(\theta)}{p(X)}=\frac{p(X / \theta) p(\theta)}{\int_{-\infty}^{\infty} p(X / \theta) p(\theta) d\theta}.$$

Учитывая, что знаменатель в этой формуле не зависит от θ, указанную выше взаимосвязь представляют в виде

$$p(\theta / X) \sim p(X / \theta) p(\theta)=\tilde{p}(X / \theta),$$

где ~ означает пропорциональность.

До конца 80-х годов прошлого века для построения байесовских оценок параметров на основе апостериорных распределений в качестве вычислительных методов использовали аналитические методы, такие как сопряжённые апостериорные распределения и т.д. Но в результате стремительного развития информационных технологий с конца 90-х годов 20 века начали распространяться новые методы, которые базируются на непосредственном моделировании выборки из апостериорных распределений. Одним из подходов к моделированию является моделирование по принципу Монте-Карло. Среди методов Монте-Карло в свою очередь выделяют итерационные методы Монте-Карло, которые базируются на основе построения Марковских цепей. К таким методам относится генерирование выборки по Гиббсу [5,6], алгоритм Метрополиса, Метрополиса – Гастингса [3, 4]. Основой моделирования выборки с помощью MCMC – метода служит построение Марковского процесса, для которого стационарное распределение переходов определяется функцией p(θ/X). Приведем общую схему MCMC – метода. Рассматривается вопрос о генерации случайной выборки -из многомерного распределения с плотностью в общем виде p(T) в многомерном пространстве с помощью методов Монте – Карло по схеме Марковских цепей (MCMC). В этих методах вводится некоторая цепь Маркова с априорным распределением $p_0(T)$ и вероятностями переходов $q_n(T_{n+1}/T_n)$ генерация выборки осуществляется следующим образом:

$$T_1 \sim p_0(T), \quad T_2 \sim q_1(T_2 / T_1), \quad T_N \sim q_{N-1}(T_{N-1} / T_N). \tag{2}$$

Заметим, что сгенерированная таким образом выборка не является совокупностью независимых случайных величин. Однако, она подходит для оценки вероятностных интегралов вида (1). Часто берут однородную цепь Маркова, когда вероятность перехода $q_n(T_n/T_{n-1})$ не зависит от момента времени n, т.е. $q_n(T_n/T_{n-1} = q(T_n/T_{n-1})$. Основной вопрос, как выбрать вероятности перехода $q(T_n/T_{n-1})$, чтобы выборка, сгенерированная по схеме (2), была выборкой из распределения p(T). Один из вариантов выбора $q(T_n/T_{n-1})$ дается алгоритмом Метрополиса-Гастингса.

Он состоит в следующем. Пусть на шаге n сгенерировано значение T_n, тогда на шаге n+1 генерируется точка T^* из некоторого предложенного распределения $r(T/T_n)$ (например, нормального). Затем вычисляется величина

$$\alpha\left(T^*, T_n\right)=\min\left(1,\ \frac{\tilde{p}\left(T^*\right) r\left(T_n / T^*\right)}{\tilde{p}\left(T_n\right) r\left(T^* / T_n\right)}\right).$$

Точка T^* принимается в качестве следующей точки T_{n+1} с вероятностью $\alpha(T^*, T_n)$. С вероятностью $1 - \alpha(T^*, T)$ принимается $T_{n+1} = T_n$. Таким образом, определена цепь Маркова с вероятностью перехода

$$q\left(T_{n+1} / T_n\right)=\begin{cases} r\left(T_{n+1} / T_n\right) \alpha\left(T_{n+1}, T_n\right),\ \text{если}\ \ T_{n+1} \neq T_n, \\ 1-r\left(T_{n+1} / T_n\right) \alpha\left(T_{n+1}, T_n\right),\ \text{если}\ \ T_{n+1}=T_n. \end{cases}$$

Алгоритм стартует из некоторого случайного значения T_0 и несколько первых итераций игнорируются для того, чтобы забыть о начальном значении. Алгоритм работает лучше всего, когда вспомогательная функция $r(T_{n+1}/Tn)$ близка к форме целевой функции $p(T)$.

В данной работе на основе байесовского подхода в машинном обучении рассмотрена задача оценивания параметров временных рядов, которая сводится к нахождению условных математических ожиданий. Условные математические ожидания являются оптимальными в среднеквадратическом смысле оценками. Численная реализация условных математических ожиданий сводится к вычислению многомерных интегралов, для вычисления которых используется $MCMC$ – метод. На основе алгоритма Метрополиса-Гастингса построены цепи Маркова, которые использованы для приближенного вычисления условного математического ожидания $MCMC$ – методом. Проведены компьютерные эксперименты, показывающие эффективность предлагаемого алгоритма оценивания.

Литература

1. Harrison, B. Deep Learning and Bayesian Methods /B Hamson // The European Physical Joumal Conferences. - 2017. P 52 – 61.

2. Jun, Zhu. Learning with Bayesian Methods /Jun Zhu, Jianfei Chen, Wenbo Hu Big, Bo Zhang. 2017. Learning with Bayesian Methods. // National Science Review 2010. - P. 133 - 157.

3. Geweke, J. 2012. Bayesian estimation of state-space models using M-H algorithm and Gibbs sampling. /J. Geweke //Comput. Statistics and Data Analysis. 2012. - №2. P. 117– 130.

4. Gholami, G.H. Fundamental Concepts of MCMC Methods Proved on Metropolis-Hastings Algorithm. / G.H. Gholami, A. Etemadi, E. Fayyazi. // International Journal of advanced research in Computer Science. 2012. - №2. P. 114– 172.

5. Geman, S. 1984. Stochastic Relaxation, Gibbs Distribution, and Bayesian Restoration of Images./ S. Geman // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1984.-№ 6. – P. 721 – 741.

6. Barber, D. Bayesian reasoning and machine learning. Cambridge University Press./D. Barber. 2019. - 666 p.