Sử dụng Bayesian Optimization trong việc lựa chọn tham số cho bài toán Machine Learning

Phạm Đức Duy

Ngày 23 tháng 5 năm 2023

Khó khăn trong Học Máy?

Lựa chọn Hyperparameter phù hợp

Khó khăn trong Học Máy?

- Lựa chọn Hyperparameter phù hợp
- Cần tối ưu hoá tham sô trong không gian rộng và nhiều chiều

Khó khăn trong Học Máy?

- Lựa chọn Hyperparameter phù hợp
- Cần tối ưu hoá tham sô trong không gian rộng và nhiều chiều
- Tốn thời gian và chi phí



Bayesian Optimization là gì?

 Một chiến lược thiết kế tuần tự cho bài toán tối ưu hóa toàn cục của các hàm black-box



Bayesian Optimization là gì?

 Một chiến lược thiết kế tuần tự cho bài toán tối ưu hóa toàn cục của các hàm black-box



 Sử dụng phân phối tiên nghiệm và hậu nghiệm để mô hình hóa hàm muc tiêu

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

trong đó A và B là hai sự kiện, P(A|B) là xác suất của A xảy ra khi biết B đã xảy ra, P(B|A) là xác suất của B xảy ra khi biết A đã xảy ra, P(A) và P(B) là xác suất tiên nghiệm của A và B.

Bayesian Optimization là gì?

 Một chiến lược thiết kế tuần tự cho bài toán tối ưu hóa toàn cục của các hàm black-box



 Sử dụng phân phối tiên nghiệm và hậu nghiệm để mô hình hóa hàm muc tiêu

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

trong đó A và B là hai sự kiện, P(A|B) là xác suất của A xảy ra khi biết B đã xảy ra, P(B|A) là xác suất của B xảy ra khi biết A đã xảy ra, P(A) và P(B) là xác suất tiên nghiệm của A và B.

 Sử dụng các hàm thu nhận (acquisition functions) để quyết định điểm truy vấn tiếp theo

Vì sao Bayesian Optimization hiệu quả?

• Giảm thiểu số lần đánh giá hàm mục tiêu

Vì sao Bayesian Optimization hiệu quả?

- Giảm thiểu số lần đánh giá hàm mục tiêu
- Thích hợp cho các bài toán có chi phí đánh giá cao, nhiễu hoặc không liên tục

Vì sao Bayesian Optimization hiệu quả?

- Giảm thiểu số lần đánh giá hàm mục tiêu
- Thích hợp cho các bài toán có chi phí đánh giá cao, nhiễu hoặc không liên tục
- Không cần biết dạng hàm của hàm mục tiêu



• Hàm mục tiêu: sử dụng để đánh giá và đo lường hiệu suất của mô hình Machine Learning với các tham số cụ thể.

- Hàm mục tiêu: sử dụng để đánh giá và đo lường hiệu suất của mô hình Machine Learning với các tham số cụ thể.
- Tham số cần điều chỉnh: Đây là tập hợp các tham số có thể được tinh chỉnh và lựa chọn trong mô hình Machine Learning. Mỗi tham số đại diện cho một khía cạnh của mô hình và có thể ảnh hưởng đến hiêu suất của nó.

- Hàm mục tiêu: sử dụng để đánh giá và đo lường hiệu suất của mô hình Machine Learning với các tham số cụ thể.
- Tham số cần điều chỉnh: Đây là tập hợp các tham số có thể được tinh chỉnh và lựa chọn trong mô hình Machine Learning. Mỗi tham số đại diện cho một khía cạnh của mô hình và có thể ảnh hưởng đến hiêu suất của nó.
- Quá trình tìm kiếm: Đây là quá trình mà Bayesian Optimization sử dụng để tìm ra tham số tối ưu. Quá trình này bao gồm việc thử nghiệm các giá trị tham số khác nhau, đánh giá hiệu suất của mô hình với các giá trị tham số đó.

- Hàm mục tiêu: sử dụng để đánh giá và đo lường hiệu suất của mô hình Machine Learning với các tham số cụ thể.
- Tham số cần điều chỉnh: Đây là tập hợp các tham số có thể được tinh chỉnh và lựa chọn trong mô hình Machine Learning. Mỗi tham số đại diện cho một khía cạnh của mô hình và có thể ảnh hưởng đến hiệu suất của nó.
- Quá trình tìm kiếm: Đây là quá trình mà Bayesian Optimization sử dụng để tìm ra tham số tối ưu. Quá trình này bao gồm việc thử nghiệm các giá trị tham số khác nhau, đánh giá hiệu suất của mô hình với các giá trị tham số đó.
- Xác định tham số tối ưu: Đây là quá trình cuối cùng của Bayesian Optimization khi tìm ra giá trị tham số tốt nhất cho mô hình Machine Learning.

Bayesian Optimization hoạt động như thế nào?

• Sử dụng 1 hàm surrogate để xấp xỉ linh hoạt giá trị thay cho hàm ban đầu

Bayesian Optimization hoạt động như thế nào?

- \bullet Sử dụng 1 hàm surrogate để xấp xỉ linh hoạt giá trị thay cho hàm ban đầu
- Đặt 1 phân phối tiên nghiệm cho hàm Surrogate f(x), sau đó tìm giá trị tại các điểm x_i cho hàm mục tiêu f(x) để cập nhật phân phối tiên nghiệm và thu được phân phối hậu nghiệm.

Bayesian Optimization hoạt động như thế nào?

- \bullet Sử dụng 1 hàm surrogate để xấp xỉ linh hoạt giá trị thay cho hàm ban đầu
- Đặt 1 phân phối tiên nghiệm cho hàm Surrogate f(x), sau đó tìm giá trị tại các điểm x_i cho hàm mục tiêu f(x) để cập nhật phân phối tiên nghiệm và thu được phân phối hậu nghiệm.
- Sử dụng một hàm acquisition function để chọn điểm tiếp theo để đánh giá hàm mục tiêu, hàm này cần cân bằng giữa việc khám phá (exploration) và khai thác (exploitation)

Chuẩn bị

• Xác định không gian tham số: Việc lựa chọn tham số cần điều chỉnh là rất quan trọng, mỗi tham số sẽ có 1 khoảng giá trị hoặc tập giá trị phù hợp. Ở đây ta có thể điều chỉnh tham số **Learning rate**. Trong 1 mô hình Machine Learning hay Deep Learning. LR có thể nằm trong khoảng từ 10^{-5} đến 0,1

Chuẩn bị

- Xác định không gian tham số: Việc lựa chọn tham số cần điều chỉnh là rất quan trọng, mỗi tham số sẽ có 1 khoảng giá trị hoặc tập giá trị phù hợp. Ở đây ta có thể điều chỉnh tham số **Learning rate**. Trong 1 mô hình Machine Learning hay Deep Learning. LR có thể nằm trong khoảng từ 10^{-5} đến 0,1
- Xác định hàm mục tiêu: Để đánh giá hiệu suất của mô hình, có thể chọn hàm mục tiêu phù hợp. Các hàm mục tiêu có thể chọn bao gồm: Accuracy, Loss, hay F1-score... tuỳ vào mô hình đang sử dụng.

Surrogate function(hàm thay thế)

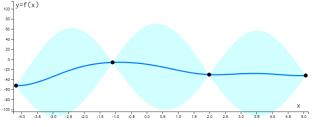
 Là 1 hàm để xấp xỉ hàm mục tiêu không khả dụng để tìm kiếm giá trị tối ưu tiềm năng

Surrogate function(hàm thay thế)

- \bullet Là 1 hàm để xấp xỉ hàm mục tiêu không khả dụng để tìm kiếm giá trị tối ưu tiềm năng
- Quá trình tìm kiếm bao gồm chọn và thử nghiệm các bộ tham số khác nhau dưa trên

Surrogate function(hàm thay thế)

- \bullet Là 1 hàm để xấp xỉ hàm mục tiêu không khả dụng để tìm kiếm giá trị tối ưu tiềm năng
- Quá trình tìm kiếm bao gồm chọn và thử nghiệm các bộ tham số khác nhau dựa trên
- Sử dụng một hàm acquisition function để chọn điểm tiếp theo để đánh giá hàm mục tiêu, hàm này cần cân bằng giữa việc khám phá (exploration) và khai thác (exploitation)



- là hàm được sử đụng để đánh giá và lựa chọn các điểm trong không gian tham số để đánh giá tiếp theo, hàm này đo lường mước độ hứa hẹn của 1 điểm thử nghiệm cho việc cải thiện hiệu suất của mô hình
- Mục tiêu của hàm là tìm ra các điểm tiềm năng có khả năng cung cấp thông tin hữu ích để cải thiện mô hình

Có nhiều loại hàm Acquisition khác nhau, tuỳ thuộc vào đặc trưng của bài toán:

Có nhiều loại hàm Acquisition khác nhau, tuỳ thuộc vào đặc trưng của bài toán:

Expected Improvement (EI): Ước tính giá trị cải thiện kỳ vọng so với điểm tốt nhất hiện tại.

Có nhiều loại hàm Acquisition khác nhau, tuỳ thuộc vào đặc trưng của bài toán:

- Expected Improvement (EI): Ước tính giá trị cải thiện kỳ vọng so với điểm tốt nhất hiện tại.
- Probability of Improvement (PI): Ước tính xác suất cải thiện so với điểm tốt nhất hiện tại.

Có nhiều loại hàm Acquisition khác nhau, tuỳ thuộc vào đặc trưng của bài toán:

- Expected Improvement (EI): Ước tính giá trị cải thiện kỳ vọng so với điểm tốt nhất hiện tại.
- Probability of Improvement (PI): Ước tính xác suất cải thiện so với điểm tốt nhất hiện tại.
- Upper Confidence Bound (UCB): Ước tính Upper Confidence Bound của giá trị kỳ vọng của một điểm dữ liệu.

Expected Improvement(EI)

Công thức của Expected Improvement (EI):

$$EI(x) = \max(0, \mu(x) - f_{\mathsf{best}} - \xi)\Phi(Z) + \sigma(x)\phi(Z)$$

Trong đó:

- $\mu(x)$ là giá trị kỳ vọng của surrogate function tại x.
- ullet $f_{
 m best}$ là giá trị tốt nhất đã được tìm thấy đến hiện tại.
- ξ là một siêu tham số để điều chỉnh tham vọng (exploitation vs exploration) của thuật toán.
- $\sigma(x)$ là độ lệch chuẩn của surrogate function tại x.
- $\Phi(Z)$ và $\phi(Z)$ là hàm phân phối tích lũy và hàm mật độ xác suất của phân phối chuẩn đối với giá trị $Z = (\mu(x) f_{\text{best}} \xi)/\sigma(x)$.

Quá trình tối ưu

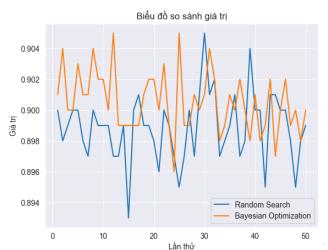
- Sử dụng Surrogate function để ước tính hàm mục tiêu và xác định xác suất tối ưu của các vị trí của tham số
- Sử dụng Acquisition function để lựa chọn vị trí tiếp theo để thử nghiệm
- Thử nghiệm mô hình với các tham số tại vị trí đã chọn
- O Cập nhật mô hình xác suất dựa trên kết quả thử nghiệm mới

Quá trình tối ưu

Quy trình tối ưu có thể dừng sau một số lượng bước cố định hoặc khi đạt được điều kiện dừng nhất định (ví dụ: đạt đủ độ chính xác mong muốn). Giá trị tham số tốt nhất được chọn là giá trị tham số tại vị trí có giá trị hàm mục tiêu tốt nhất đã được tìm thấy trong quá trình tối ưu.

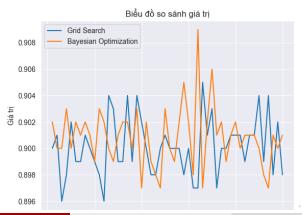
So sánh với các phương pháp khác

Random search: Bayesian đạt hiệu quả hơn đáng trong thấy với bộ dữ liệu có nhiều chiều



So sánh với các phương pháp khác

Grid search: Khó có thể đánh giá độ hiệu quả của 2 thuật toán nếu không giới hạn số lần thử. Tuy nhiên nếu giới hạn số lần thử của thuật toán, Bayesian Optimization chiếm ưu thế vì tìm ra tham số tối ưu sớm hơn Grid Search.



Các ứng dụng của Bayesian Optimization trong học máy

• Tối ưu hóa các tham số siêu (hyperparameters) của các mô hình phức tạp như mạng nơ-ron

Các ứng dụng của Bayesian Optimization trong học máy

- Tối ưu hóa các tham số siêu (hyperparameters) của các mô hình phức tạp như mạng nơ-ron
- Tìm kiếm kiến trúc mô hình tốt nhất

Các ứng dụng của Bayesian Optimization trong học máy

- Tối ưu hóa các tham số siêu (hyperparameters) của các mô hình phức tạp như mạng nơ-ron
- Tìm kiếm kiến trúc mô hình tốt nhất
- Tối ưu hóa các chức năng không khả vi (non-differentiable) như độ chính xác hoặc F1-score...

Tài liệu tham khảo

- Dịnh nghĩa và ý tưởng chọn dataset cho code được tham khảo từ: "Fast Bayesian Optimization of Machine Learning Hyperparameters on Large Datasets" https://arxiv.org/abs/1605.07079
- ② Bài báo cáo còn sử dụng 1 số hình ảnh từ các nguồn bên ngoài như: https://thetalog.com/thetaflow/ bayesian-optimization-visualizer/