

Permasalahan Multi-Armed Bandit dengan Piecewise-Stationary Bernoulli Arms = Multi-Armed Bandit Problem with Piecewise-Stationary Bernoulli Arms

Faathir Chikal Asyuraa, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509599&lokasi=lokal>

Abstrak

Permasalahan Multi-Armed Bandit adalah permasalahan dalam reinforcement learning yang berfokus pada rancangan eksperimen, diberikan sebuah himpunan opsi yang disebut arms yang dapat dipilih berkali-kali, bagaimana cara menyeimbangkan antara mengeksplorasi arm yang ada untuk mengumpulkan informasi atau mengeksploitasi arm yang terlihat terbaik untuk memaksimalkan keuntungan. Oleh karena itu, Multi-Armed Bandit menjadi alternatif yang lebih dinamis dari percobaan acak. Contoh dari aplikasi Multi-Armed Bandit adalah menentukan artwork film yang harus ditunjukkan untuk menarik pengunjung untuk menonton film tersebut. Distribusi Bernoulli dengan parameter θ dipilih untuk memodelkan respons dari pengunjung setelah melihat artwork film. Kondisi tidak stasioner pada θ dapat diimplementasikan untuk mengakomodasi periode keunggulan berbeda dalam artwork film. Kondisi tidak stasioner pada studi ini dimodelkan melalui piecewise-stationary, yaitu θ dapat berubah nilai, namun tetap konstan di setiap periode yang didefinisikan. Pada penelitian ini, digunakan beberapa policy seperti Epsilon Greedy, SoftMax, Upper Confidence Bounds, Thompson Sampling, Sliding Window Upper Confidence Bounds, Discounted Upper Confidence Bounds, dan juga Discounted Thompson Sampling, untuk menangani permasalahan Multi-Armed Bandit dengan Piecewise-Stationary Bernoulli Arms. Simulasi dilakukan pada kondisi yang berbeda-beda untuk menguji performa policy tersebut dalam berbagai kondisi yang ada. Berdasarkan simulasi tersebut, Discounted Thompson Sampling policy menunjukkan performa yang sangat baik dalam menangani kondisi stasioner maupun piecewise-stationary.

.....The Multi-Armed Bandit problem is a problem in reinforcement learning that focuses on how to design an experiment, given a set of options called arms that could be tried many times, how to balance between exploring the available arms to gather information or exploiting the seemingly best arm to maximize profit. Because of this, Multi-Armed Bandit has gained its popularity as a more dynamic approach to a randomized trial. An example of Multi-Armed Bandit is in determining recommending a film artwork to show to a visitor. Bernoulli distribution with parameter θ is chosen to model the respons of the visitor whether they watch the film or not. Non-stationary condition on θ can be implemented to accommodate various trends in film artworks, the non-stationary condition in this study is modeled through Piecewise-Stationary. In this study, several policies are used, such as Epsilon Greedy, SoftMax, Upper Confidence Bounds, Thompson Sampling, Sliding Window Upper Confidence Bounds, Discounted Upper Confidence Bounds, and Discounted Thompson Sampling, in handling Multi-Armed Bandit with Piecewise-Stationary Bernoulli Arms. Multiple simulations have been done to empirically evaluate the performance of the policies. Based on the simulation, Discounted Thompson Sampling policy shows a remarkable performance in tackling stationary and piecewise-stationary condition.