

ABSTRAK

Perilaku pencarian sumber daya yang tersebar secara acak adalah permasalahan ekologis yang umum dimodelkan dalam berbagai implementasi bidang ilmu pengetahuan. Berbagai penelitian mensolusikan permasalahan ini dengan menggunakan strategi pencarian acak seperti *Random Waypoint* dan *Brownian Motion*, namun pencarian acak heuristik *Lévy Flight* disebut lebih optimal. Walaupun disebut optimal, *Lévy Flight* memiliki kelemahan karena nilai heuristik yang tepat dibutuhkan supaya dapat mencari target dengan efektif. Untuk mengatasi kelemahan ini, kami mengusulkan algoritma pencarian yang adaptif menggunakan *Reinforcement Learning* berjenis *Temporal Difference* agar dapat menyesuaikan lintasan pencarian dengan pola persebaran sumber daya. Penelitian ini membandingkan performa variasi strategi eksplorasi-eksploitasi menggunakan algoritma *Q-Learning* untuk menemukan pemilihan lintasan pencarian yang optimal. Menggunakan strategi *Epsilon Greedy*, *Upper Confidence Bound*, dan *Posterior Sampling*, kami mengevaluasi performa mereka dengan *Lévy Flight* dalam berbagai pola persebaran sumber daya seperti *Homogeneous Poisson Process* dan *Thomas Cluster Process* di lingkup area dua dimensi. Agen belajar hanya dengan mengobservasi *reward* yang didapatkan dari interaksi sebelumnya pada lingkungan dengan mengandalkan perhitungan jumlah langkah lurus sebagai representasi *state*. Hasil penelitian ini menunjukkan keunggulan algoritma pencarian yang belajar dengan *Upper Confidence Bound* dan *Posterior Sampling* yang menghasilkan akumulasi *reward* lebih tinggi, serta panjang lintasan yang lebih sesuai terhadap pola persebaran sumber daya di lingkungan dibandingkan dengan *Lévy Flight* dan *Epsilon Greedy*.

Kata kunci: strategi pencarian sumber daya, *lévy flight*, *q-learning*, strategi eksplorasi-eksploitasi

ABSTRACT

Foraging behavior is an ecological phenomenon commonly modeled across some fields. A number of studies have conducted the solution using variety of random search strategies such as Random Waypoint and Brownian Motion, with Lévy Flight heuristic search generally being regarded as more optimal. Be that as it may, Lévy Flight possess a weakness around the need of the right heuristic value to be able to effectively search for the targets. To address this limitation, we propose an adaptive search algorithm using Temporal Difference Reinforcement Learning that dynamically adjust the search trajectory based on the resource distribution pattern. This research compares Epsilon Greedy, Upper Confidence Bound, and Posterior Sampling exploration-exploitation strategy integrated into Q-Learning algorithm to seek the optimal search trajectory selection. We evaluate their performance against Lévy Flight in various resource distribution patterns such as Homogeneous Poisson Process and Thomas Cluster Process in a two-dimensional area. The learning agent observes only the reward obtained from past interactions in the environment and learns by utilizing the consecutive number of straight steps as its representation of state. Our results show the superiority of the learning search algorithm with Upper Confidence Bound and Posterior Sampling against Lévy Flight and Epsilon Greedy strategy, yielding higher total rewards obtained as well as its suitability of trajectory length to the resource distribution pattern in the environment.

Keywords: foraging strategy, lévy flight, q-learning, exploration-exploitation strategy