International Journal of Swarm Intelligence Research
Volume 16 • Issue 1 • January-December 2025

Adaptive English Translation Parameter Tuning via Particle Swarm Optimization and Attention Mechanism

Zhigao Wei
https://orcid.org/0009-0002-8178-3336
Xinxiang University, China

ABSTRACT

This paper proposes an English Translation Model Parameter Adaptive Tuning Method Integrating Particle Swarm Optimization (PSO) and Attention Mechanism. We encode key Neural Machine Translation (NMT) hyperparameters as particles and employ a specialized fitness function that rewards both conventional translation metrics and attention alignment quality. Furthermore, a dynamic feedback multiplier modulates PSO velocity updates in real time, incentivizing hyperparameter adjustments when attention scores deviate from predefined thresholds. Experimental evaluations on both IWSLT En–De and WMT En–De datasets demonstrate the effectiveness of this method, yielding improvements in BLEU and ROUGE-L scores relative to standard baselines. Ablation studies further confirm that both the attention alignment term in the fitness function and the dynamic attention feedback are indispensable for achieving superior translation quality and more coherent attention patterns.

KEYWORDS

English Translation, Adaptive Tuning, Particle Swarm Optimization, Neural Machine Translation

INTRODUCTION

In recent years, the rapid development of deep learning in the field of natural language processing has significantly advanced machine translation techniques (Li, 2018). Neural machine translation (NMT) models, characterized by their end-to-end learning paradigms, have demonstrated impressive performance on multilingual translation tasks. A typical NMT model is based on a sequence-to-sequence (Seq2Seq) architecture and often incorporates attention mechanisms to enhance translation quality and fluency (Gehring, Auli, Grangier, Yarats, et al., 2017). However, as model size and complexity continue to grow, the number of hyperparameters (e.g., learning rate, regularization coefficients, number of layers, and attention heads) also increases. Optimal configurations for these hyperparameters can vary widely depending on the specific dataset and task settings.

To achieve the best translation performance in diverse linguistic contexts and across different data scales, adaptive tuning of model parameters has become a critical research topic. Traditional hyperparameter search methods, such as grid search or random search, can be computationally expensive and time-consuming, especially when training deep NMT models. Consequently, there is a pressing need for efficient search strategies that can navigate large hyperparameter spaces and converge to optimal configurations in a timely manner.

DOI: 10.4018/IJSIR.372900

Although some researchers have explored the use of evolutionary or swarm intelligence algorithms in natural language processing tasks, their application to machine translation remains relatively limited. Deep NMT models typically involve iterative training processes that are computationally intensive; simple brute-force or random searches can become prohibitively expensive in terms of hardware and time requirements. Moreover, state-of-the-art hyperparameter optimization methods, such as Bayesian optimization or gradient-based searches, depend heavily on a thorough understanding of the problem space or gradient information, which may be insufficient for highly complex translation data. Additionally, given the importance of attention mechanisms in NMT, it is vital to retain or enhance the benefits of attention while also employing swarm intelligence techniques to efficiently explore and adjust attention-related parameters.

Swarm intelligence is a branch of artificial intelligence inspired by the collective behavior of social organisms, such as ant colonies, flocks of birds, and schools of fish. Central to this paradigm is the concept that decentralized and self-organizing agents can cooperate to solve complex problems, often outperforming singular, centrally controlled strategies. Commonly studied algorithms within swarm intelligence include ant colony optimization, particle swarm optimization (PSO), and artificial bee colony, each leveraging a simple set of agent-level rules and local interactions to collectively explore a problem space. One of the most influential techniques in this domain is PSO. PSO models a population of candidate solutions (particles) that navigate a high-dimensional search space by combining aspects of local (individual) experience with information gleaned from global or neighborhood best performers. This combination of exploitation (following known good solutions) and exploration (probabilistic, random movements) fosters a balance that helps avoid premature convergence, making PSO well-suited for non-convex and multimodal optimization problems. Swarm intelligence has proven effective for hyperparameter tuning in NMT and other deep learning tasks due to its adaptability and collective search dynamics. Unlike grid or random search, swarm-based methods quickly converge on promising regions of large, high-dimensional spaces while lowering the risk of getting stuck in suboptimal solutions. By incorporating task-specific signals, such as attention alignment for NMT, swarm algorithms maintain contextual awareness and ultimately guide the search toward hyperparameter configurations that improve translation quality. This makes swarm intelligence particularly appealing for large-scale and iterative tuning in modern deep learning.

This paper proposes an English translation model parameter adaptive tuning method integrating PSO and attention mechanisms. The main framework is as follows. First, I encoded critical hyperparameters of the NMT model and generated an initial population of particles within the hyperparameter space. Each particle represented a candidate solution. After each round of model training, I employed an attention-based evaluation strategy to assess translation quality, incorporating metrics, like bilingual evaluation understudy (BLEU) and recall-oriented understudy for gisting evaluation (ROUGE), as well as an assessment of attention distributions. This evaluation forms a comprehensive fitness function that captures both overall translation accuracy and the contribution of attention alignments. Next, I used the global best and personal best solutions within the swarm to iteratively update the velocities and positions of the particles. By leveraging the swarm's collective and individual search behaviors, the PSO algorithm refined the hyperparameter candidates, steering them toward optimal solutions in a cost-effective manner. Furthermore, a dynamic feedback mechanism from the attention module was introduced, providing real-time information on how attention contributes to translation performance. This feedback allowed for fine-grained adjustments of hyperparameters, thereby enhancing the effectiveness and stability of the training process. Finally, the algorithm stopped either when it reached a maximum number of iterations or when there was no significant improvement in translation metrics over a predefined number of rounds, resulting in an optimized hyperparameter configuration for English translation tasks. The contributions are listed as follows.

Existing NMT hyperparameter tuning methods face a multifaceted set of challenges due to the increasing depth and complexity of contemporary models. First, exhaustive approaches, like grid search, become infeasible as the number of hyperparameters grows exponentially. Second, random

Related Content

An Improved Sparrow Approach
Wenfang Chen (2022). *International Journal of Swarm Intelligence Research (pp. 1-17).*
www.irma-international.org/article/an-improved-sparrow-approach/313667

Securing Healthcare Systems Integrated With IoT: Fundamentals, Applications, and Future Trends
S. Boopathi (2023). *Dynamics of Swarm Intelligence Health Analysis for the Next Generation (pp. 186-209).*
www.irma-international.org/chapter/securing-healthcare-systems-integrated-with-iot/326195

Path Relinking with Multi-Start Tabu Search for the Quadratic Assignment Problem
Tabitha Jamesand Cesar Rego (2011). *International Journal of Swarm Intelligence Research (pp. 52-70).*
www.irma-international.org/article/path-relinking-multi-start-tabu/55320

Harvesting Tomorrow: Transformative Applications of Swarm Robotics in Agriculture
Sima Das (2024). *Bio-inspired Swarm Robotics and Control: Algorithms, Mechanisms, and Strategies (pp. 193-210).*
www.irma-international.org/chapter/harvesting-tomorrow/345314

A Hybrid Grey Wolves Optimizer and Convolutional Neural Network for Pollen Grain Recognition
Hanane Menad, Farah Ben-naoumand Abdelmalek Amine (2020). *International Journal of Swarm Intelligence Research (pp. 49-71).*
www.irma-international.org/article/a-hybrid-grey-wolves-optimizer-and-convolutional-neural-network-for-pollen-grain-recognition/255709