

INTISARI

Q-LEARNING UNTUK PEMBENTUKAN TREE DALAM SENTENCE GENERATION BAHASA INDONESIA DENGAN ALGORITME TREE LONG SHORT-TERM MEMORY

Oleh

Rochana Prih Hastuti

17/418664/PA/05448

Algoritme *Tree-LSTM* adalah varian dari *LSTM* yang mengakomodasi pemrosesan struktur *tree*. Hal ini memungkinkan untuk melakukan ekstraksi informasi di luar pola sekuens linier biasa. Namun, penggunaan *Tree-LSTM* pada permasalahan *text generation*, yang mendasari pengembangan teknologi *conversational agent* misalnya *chatbot*, memerlukan bantuan parser eksternal pada setiap iterasi pembangkitan. Pembuatan parser yang baik memerlukan representasi fitur yang kompleks dan sangat bergantung pada *grammar* dari corpus pembentuk. Jumlah corpus yang terbatas, menimbulkan kesulitan proses *text generation* karena keterbatasan kosa kata.

Penelitian ini mengusulkan penggunaan *Reinforcement Learning* dalam pembentukan *constituency tree* untuk menjembatani proses *sentence generation* dalam model *Tree-LSTM*. Proses penyusunan *tree* dimodelkan sebagai *markov decision process* yang terdiri dari himpunan *state* yakni vektor *word embedding*, dan himpunan *action* yakni $\{Shift, Reduce\}$. Model *Deep Q-Network* sebagai aproksimator pada algoritme *Q-Learning* dilatih untuk memperoleh bobot yang optimal dalam merepresentasikan *Q-value function*.

Dari hasil pengujian pada 205 kalimat tes, model kombinasi *Tree-LSTM* dan *Q-Learning* memperoleh rata-rata *perplexity* 9,6, lebih baik dari model *sentence generation* dengan bracketing *Shift All*. Model yang diusulkan memperoleh rata-rata ranking terbaik pada evaluasi responden melalui kuesioner, yakni 1,71, yang didukung dengan signifikansi pada hasil analisis *friedman test* dengan *p-value* $3,78.10^{-44}$. Hasil tersebut menunjukkan bahwa kelima responden secara umum memberikan penilaian yang sama untuk model kombinasi *Tree-LSTM* dan *Q-Learning*, yakni model mengungguli dua model tanpa *grammar* lain, *Shift All* dan *Reduce All*.

Kata kunci: *Tree-LSTM, Q-Learning, Deep Q-Network, Text Generation*

ABSTRACT

Q-LEARNING FOR TREE FORMATION IN SENTENCE GENERATION OF BAHASA INDONESIA USING TREE LONG SHORT-TERM MEMORY ALGORITHM

By

Rochana Prih Hastuti

17/418664/PA/05448

Tree-LSTM algorithm is a variant of LSTM that accommodates tree structure processing. This makes it possible to extract information outside the linear sequence pattern. However, the use of Tree-LSTM in text generation problems, which underlies the development of conversational agent technology such as chatbot, requires the help of an external parser at each generation iteration. Developing a good parser demands the representation of complex features and relies heavily on the grammar of the corpus. The limited number of corpus causes difficulties in the text generation process, due to the limited vocabularies.

This research proposes the use of a Reinforcement Learning algorithm in the formation of constituency trees, which link the sentence generation process in the Tree-LSTM model. The tree production process is modeled as a Markov's decision process, where a set of states consists of word embedding vectors, and a set of actions of Shift, Reduce. The Deep Q-Network model as an approximator of the Q-Learning algorithm is trained to obtain optimal weights in representing the Q-value function.

From the test results on 205 test sentences, the combined Tree-LSTM and Q-Learning model achieved an average perplexity of 9.6, better than the sentence generation model with Shift-All bracketing. The proposed model obtained the best average ranking of 1.71 on the human evaluation through a questionnaire, which is supported by the significance of the results of the Friedman test analysis with $p\text{-value } 3,78.10^{-44}$. These results indicate that the five respondents generally gave the same assessment for the combination of the Tree-LSTM and Q-Learning models, which outperformed two other nongrammar models, i.e. Shift All and Reduce All.

Keywords: *Tree-LSTM, Q-Learning, Deep Q-Network, Text Generation*