

INTISARI

PREDIKSI SEMANTIC DEPENDENCY GRAPH JAWABAN PERTANYAAN GEOSPASIAL DENGAN HYBRID GRAPH LEARNING DAN GRAPH-PATTERN ASSOCIATION RULE

Oleh

IRKHAM HUDA

21/484728/SPA/00801

Semantic dependency graph (SDG) dapat mengungkap makna dan hubungan antar kata dalam kalimat pertanyaan geospasial dengan lebih baik untuk generalisasi tipe pertanyaan yang lebih luas dalam sistem *geospatial question-answering* (GeoQA). Agar SDG lebih akurat perlu peningkatan model *semantic dependency parsing* (SDP). SDP terbaik menggunakan framework *dynamic graph learning* berbasis GNN dapat ditingkatkan akurasi dengan perbaikan metode *graph structure learning* (GSL) untuk menghasilkan *adjacency matrix*, dengan mengubah GSL berbasis *multilayer perceptrons* dan *biaffine attention* dengan GSL adaptasi dari model DGSLN yang lebih memenuhi tiga prinsip *graph* dalam situasi *real-life* yaitu *homophily*, *sparsity*, dan *degree distribution*. Selain itu, pada tahap pembelajaran representasi graph, digunakan varian GNN yang lebih kuat, yaitu GIN dan GATv2 dibanding varian awalnya (GCN dan GAT). Model yang diusulkan khususnya yang terbaik menggunakan varian GIN, menunjukkan peningkatan signifikan dibandingkan model sebelumnya pada dataset Bahasa Inggris, dengan rata-rata skor F1 sebesar 93.77% pada data *in-domain* dan 92.27% pada data *out-of-domain*, meskipun kecepatan parsing-nya lebih lambat. Penggunaan *node sampling* dan perbaikan teknik *regularization* pada fungsi *hybrid loss* berpeluang untuk meningkatkan kecepatan dan akurasi model SDP kedepannya.

Salah satu tantangan dalam sistem GeoQA adalah untuk memprediksi pola jawaban yang menyerupai jawaban manusia, maka digunakan representasi SDG menggunakan metode *graph-pattern association rules* (GPAR) untuk mengungkap pola tersebut. Penelitian ini menerapkan *semantic encoding*, SDP dengan konfigurasi terbaik, dan *graph structure transformation* untuk membangun SDG dari GeoQA dengan format yang sesuai untuk proses *frequent pattern mining* (*FPMiner*) dan *rule generation* (*RuleGen*) dalam GPAR. *FPMiner* mendeteksi lebih banyak *pattern* daripada *baseline* tetapi lebih lambat, maka dipercepat dengan paralelisasi (*FPMiner_{par}*). Kombinasi *FPMiner_{par}* dan *RuleGen* memperoleh akurasi tertinggi (0.76), sedangkan *FPMiner_{par}* dengan *RepRuleGen* memperoleh nilai *confidence* tertinggi (0.72). *Rules* yang dihasilkan dapat menangkap secara efektif karakteristik pola graph GeoQA, terutama

untuk pertanyaan yang diawali dengan “*where*”, “*what*”, “*which*”, dan “*is/are*”, serta berhasil memprediksi pola SDG dari jawaban. Perbaikan metode *semantic encoding* dan pendekatan pendekatan *Large Language Models*(LLM) berpeluang meningkatkan akurasi prediksi kedepannya.

Kata kunci: Graph neural networks, Graph structure learning, Semantic Dependency Parsing, Graph Pattern Association Rules.

ABSTRACT

SEMANTIC DEPENDENCY GRAPH PREDICTION OF GEOSPATIAL QUESTION ANSWERS WITH HYBRID GRAPH LEARNING AND GRAPH-PATTERN ASSOCIATION RULE

By

IRKHAM HUDA

21/484728/SPA/00801

Semantic dependency graph (SDG) can uncover the meaning and inter-word relationships in geospatial question sentences more effectively, enabling broader generalization of question types in geospatial question-answering (GeoQA) systems. To improve SDG accuracy, enhancements to the *semantic dependency parsing* (SDP) model are necessary. The best-performing SDP—which uses a dynamic graph learning framework based on GNNs—can further boost its accuracy by refining the graph structure learning (GSL) methods for generating the adjacency matrix, replacing the GSL based on multilayer perceptrons and biaffine attention with a GSL adapted from the DGSLN model, which more fully satisfies the three graph principles in *real-life* scenarios: homophily, sparsity, and degree distribution. Moreover, at the graph representation learning stage, stronger GNN variants—namely GIN and GATv2—are employed instead of the original variants (GCN and GAT). The proposed model, particularly the variant using GIN, demonstrates significant improvements over previous models on the English dataset, achieving an average F1 score of 93.77% on in-domain data and 92.27% on out-of-domain data, albeit with slower parsing speed. The use of node sampling and enhanced regularization techniques in the hybrid loss function holds promise for further improving the speed and accuracy of the SDP model in the future.

One of the challenges in GeoQA systems is predicting answer patterns that resemble human responses; thus, SDG representations are employed using the graph-pattern association rules (GPAR) method to uncover these patterns. This study applies semantic encoding, the optimally configured SDP, and graph structure transformation to construct SDGs from GeoQA in a format suitable for the frequent pattern mining (FPMiner) and rule generation (RuleGen) processes within GPAR. FPMiner detects more patterns than the baseline but is slower, so it is accelerated through parallelization (FPMiner_{par}). The combination of FPMiner_{par} and RuleGen achieves the highest accuracy (0.76), while FPMiner_{par} with RepRuleGen attains the highest confidence (0.72). The rules generated effectively capture the characteristics of GeoQA graph

patterns, particularly for questions beginning with “where,” “what”, “which”, and “is/are” and successfully predict the SDG patterns of the answers. Improvements in semantic encoding methods and Large Language Model (LLM) approaches have the potential to further enhance predictive accuracy in future work.

Keywords: Graph neural networks, Graph structure learning, Semantic Dependency Parsing, Graph Pattern Association Rules.