[BIOINFO 2025]

LLM-Integrated Representative Path Selection for Context-Aware Drug Repurposing on Biomedical Knowledge Graphs

Haerin-Song[1], Dongmin Bang[2], Bonil Koo[2], Sun Kim,[1,2] and Sangseon Lee[3]

[1]*Interdisciplinary Program in Artificial Intelligence, Seoul National University*
[2]*Interdisciplinary Program in Bioinformatics, Seoul National University*
[3]*Department of Artificial Intelligence, Inha University*
**Corresponding author: ss.lee@inha.ac.kr*

Drug repurposing, aiming to identify novel drug-disease associations by integrating biomedical knowledge, is hindered by challenges in modeling complex multi-hop relationships in knowledge graphs. We propose **DrugCORpath**, a novel approach integrating biomedical knowledge graphs with pretrained biomedical large language models (LLMs). Unlike traditional methods that learn isolated node representations, LLM integration allows capturing of biological contexts by embedding path sentences derived from multi-hop paths connecting drug and disease entities. Each path is converted into biological path sentences reflecting plausible mechanism of action and then embedded to capture rich semantic relationships among entities. We further employ a data-driven selective filtering using K-Means clustering and distance metric to identify meaningful paths, while eliminating redundancy and noise. Experiments on datasets show our method outperforming existing graph-based, LLM-based, and path-based baselines in drug repurposing, achieving up to 4.9% higher accuracy than the prior SOTA. Further analysis confirms that selective path filtering reduces noise and enhances biological diversity. Case studies validate the clinical relevance of the selected paths, ensuring improved interpretability. This collectively underscores the method's potential for interpretable, biologically plausible drug repurposing applications.