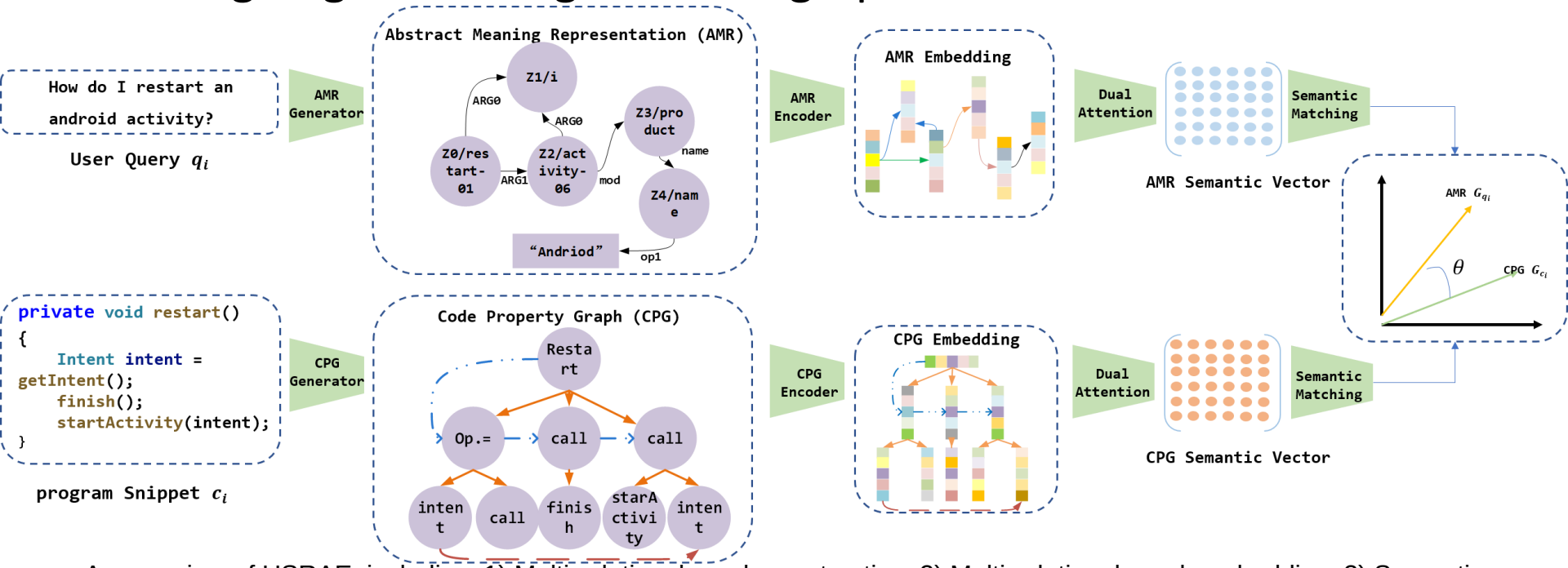
Semantic Similarity-based Program Retrieval: A Multi-relational Graph Perspective

Qianwen GOU, Yunwei DONG,
Yujiao WU, Qiao KE

Frontiers of Computer Science, DOI: 10.1007/s11704-023-2678-8

Problems & Ideas

- Problems of deep learning-based program retrieval:
  - The complementary semantic features described between different program views are often overlooked.
  - A semantic gap between natural languages and programming language.
- Ideas: an idea that frames program retrieval as a multi-relational graph similarity problem, with dual-level attention assigning node weights in the graphs.



An overview of USRAE, including: 1) Multi-relational graph construction; 2) Multi-relational graph embedding; 3) Semantic similarity calculation.

Main Contributions

- Contributions:
  - Performs program retrieval as a multi-relational graph similarity problem, employing the multi-relational graph to represent nuanced syntax and semantic features of queries and programs.
  - We address query ambiguity by abstracting them into Abstract Meaning Representation (AMR).
  - A multi-relational graph neural network based on a dual-level attention mechanism, including intra-relation and inter-relation level.

Table1. Effectiveness of USRASE in terms of SR@k and MRR metrics on JB-Java dataset (Best scores are in bold).

Models	SR@1	SR@5	SR@10	MRR
UNIF	0.435	0.652	0.756	0.493
DeepCS	0.528	0.692	0.771	0.541
CARLCS	0. 604	0.752	0.803	0.600
TABCS	0.651	0.821	0.835	0.669
DGMS	0.721	0.953	0.964	0.811
USRAE	**0.813**	**0.963**	**0.982**	**0.877**

Table 2. Effectiveness of USRAE in terms of SR@k and MRR metrics on CSN-Python dataset (Best scores are in bold)

Models	SR@1	SR@5	SR@10	MRR
UNIF	0.416	0.645	0.747	0.472
DeepCS	0.471	0.653	0.753	0.502
CARLCS	0.585	0.693	0.752	0.573
TABCS	0.662	0.721	0.792	0.628
DGMS	0.778	0.819	0.897	0.662
USRAE	**0.801**	**0.959**	**0.980**	**0.862**