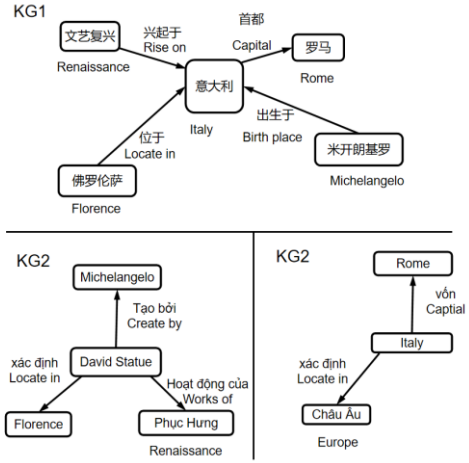
Semantic-Aware Entity Alignment for Low Resource Language Knowledge Graph

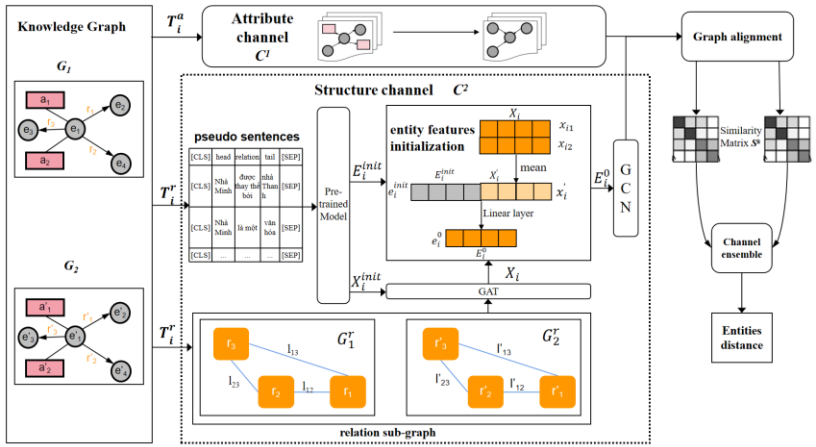
Junfei TANG, Ran SONG, Yuxin HUANG, Shengxiang GAO, Zhengtao YU

Frontiers of Computer Science, DOI: 10.1007/s11704-023-2542-x

Problems & Ideas

- Problems of low resource languages entity alignment:
  - The trained model performance absolutely depends on the amount of training data and the structure of the graph. But low resource languages often lack sufficient data to fit parameters.
  - Semantic relation as important potential information that always being ignored.
- Ideas: We proposed a Semantic-aware Graph Neural Network (SGNN) for Entity Alignment model named SGNN. Using sub-graph and pseudo sentences to get semantic relation in KGz.





Left: Without semantic information, the model may be wrong; Right: The framework of the whole model (SGNN).

Main Contributions

- Contributions:
  - We constructed a set of low resource language KG datasets between Thai, Chinese and Vietnamese.
  - We proposed an EA model with relation sub-graph to enrich the graph structure information by aggregate semantic representation with less KGs resource. And we also designed a pseudo sentence module to mine potential semantic information between entity and relation in KG effectively.
  - We evaluate our model both on three low resource languages EA datasets we built and three EA public datasets. Experimental results show that SGNN integrates relation semantics properly, and our approach significantly outperforms state-of-the-art methods for entity alignment.

model	ZH-VI		ZH-TH		TH-VI	
	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10
GCN-Align(2018) [32]	17.12	45.21	20.00	46.51	20.06	51.20
HGCN(2019) [38]	13.67	22.33	13.89	20.69	11.77	18.74
RDGCN(2019) [37]	7.12	12.69	4.97	13.60	8.80	12.29
MuGNN(2019) [33]	20.83	67.52	28.74	73.14	23.38	69.79
AttrGNN(2020) [36]	41.11	**71.75**	51.62	76.38	47.49	71.71
RNM(2021) [40]	18.83	24.14	15.94	20.91	10.91	15.09
KE-GCN(2021) [41]	43.11	67.02	50.94	74.54	46.80	72.657
SGNN	**45.17**	71.28	**53.49**	**77.46**	**51.20**	**73.14**

The performance for entity alignment of all compared approaches on the low resource language evaluation datasets. And we observe that our approach gives the best performance across all metrics and datasets, except for Hits@10 on ZH-VI where the performance of our approach is second to AttrGNN with a marginally lower score (71.28 vs. 71.75).