Evaluating the Effect of Semantic Enrichment on Entity Embeddings of IoT Knowledge Graphs

Roderick van der Weerdt, Victor de Boer, Laura Daniele, Ronald Siebes and Frank van Harmelen

May 28th, 2023

Introduction

Setting the scene

- Ongoing effort to collect IoT data
  - For example in IoT Knowledge Graphs
- There are benefits to learning over KGs
- Can we learn directly over the IoT KG?
  - Or do IoT KGs require changes?

Introduction

Setting the scene

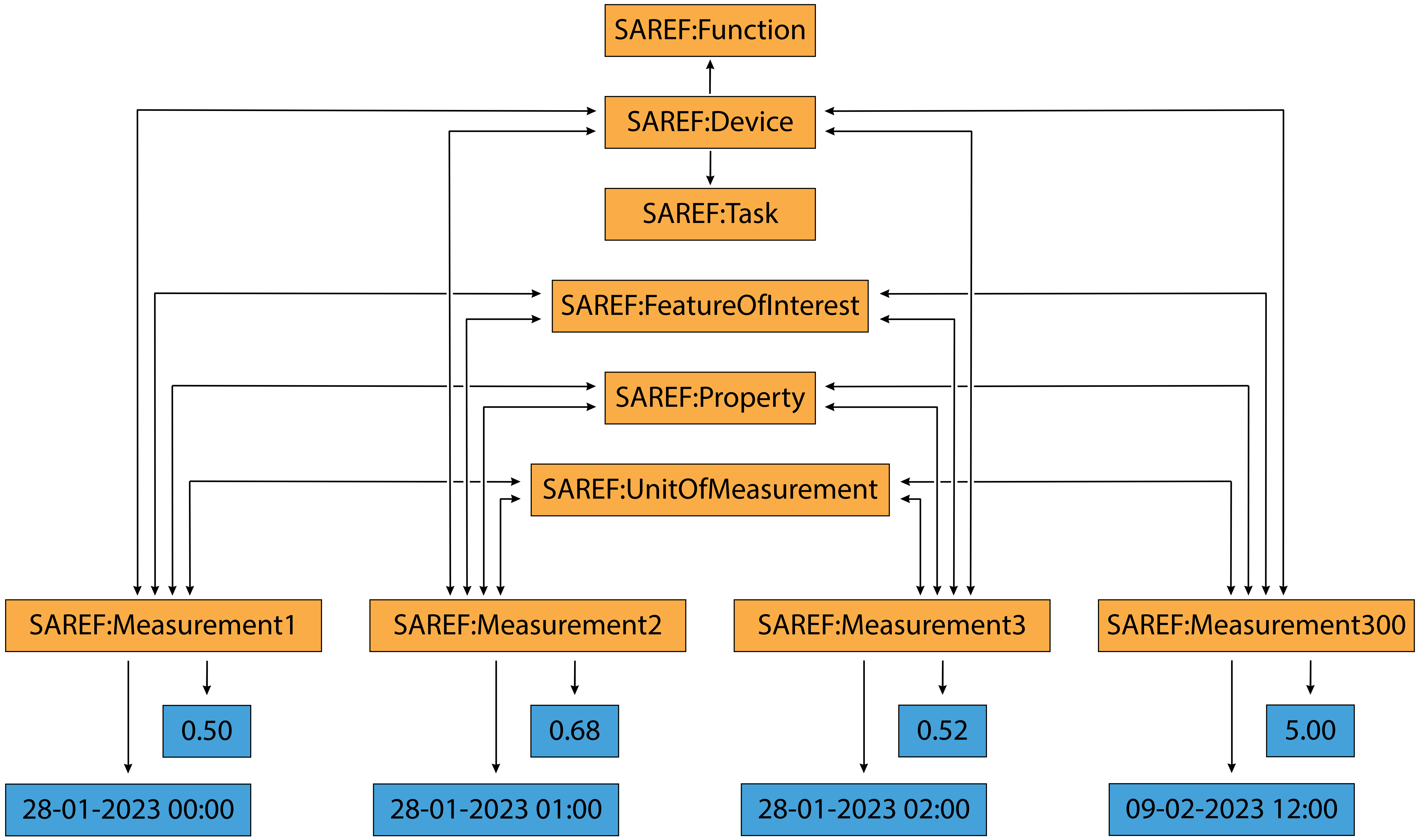
- Ongoing effort to collect IoT data
  - For example in IoT Knowledge Graphs
- There are benefits to learning over KGs
- Can we learn directly over the IoT KG?
  - Or do IoT KGs require changes?

What is the effect of semantically enriching an IoT KG, based on the quality of entity embeddings learned from it?

Introduction

What is an IoT KG?

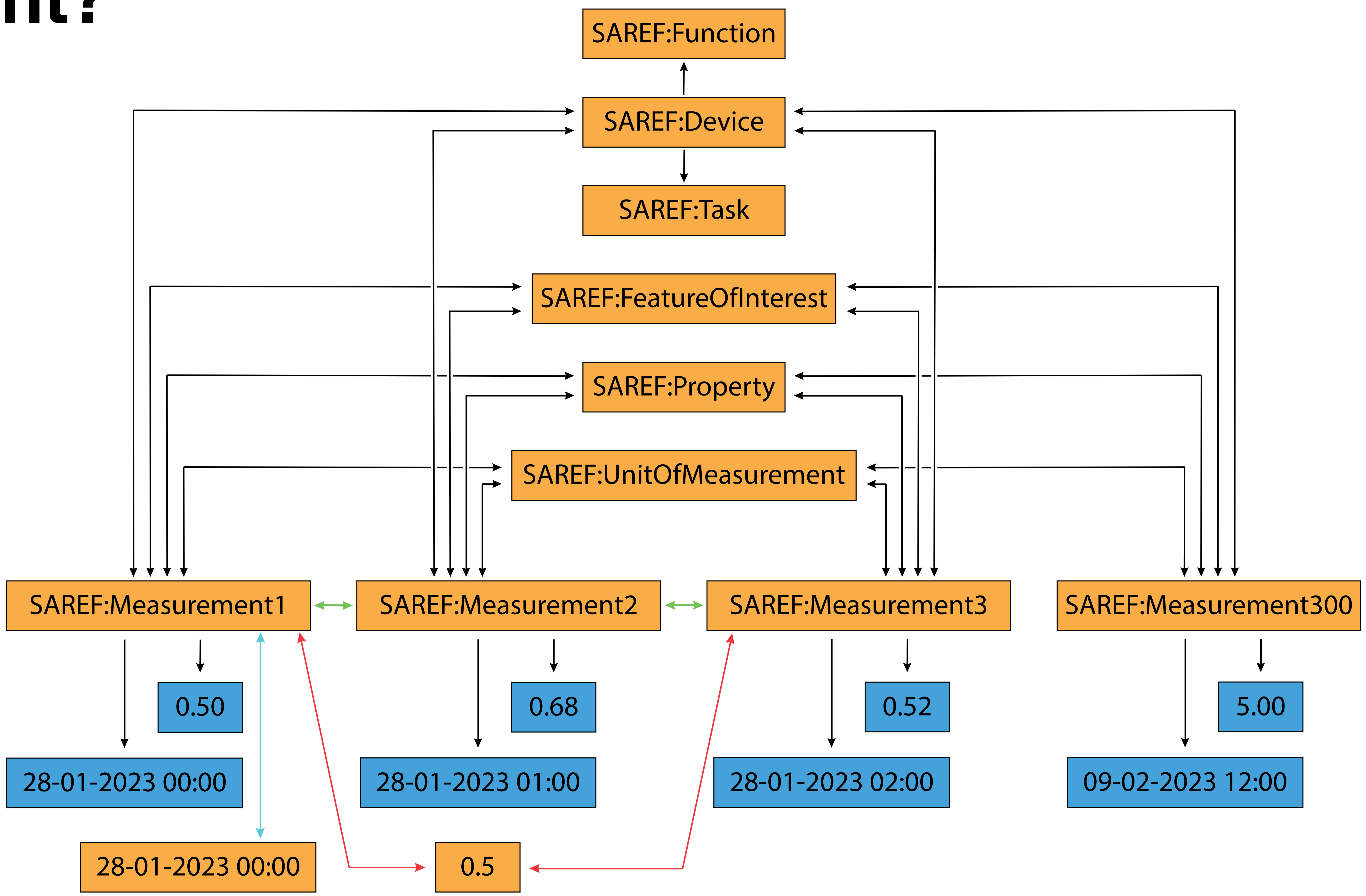
- Represent IoT measurement data
- "Wide" Graph
- *Many measurements*



Introduction

What is Semantic Enrichment?

- Making implicit information explicitly available



Introduction

What are Entity Embeddings?

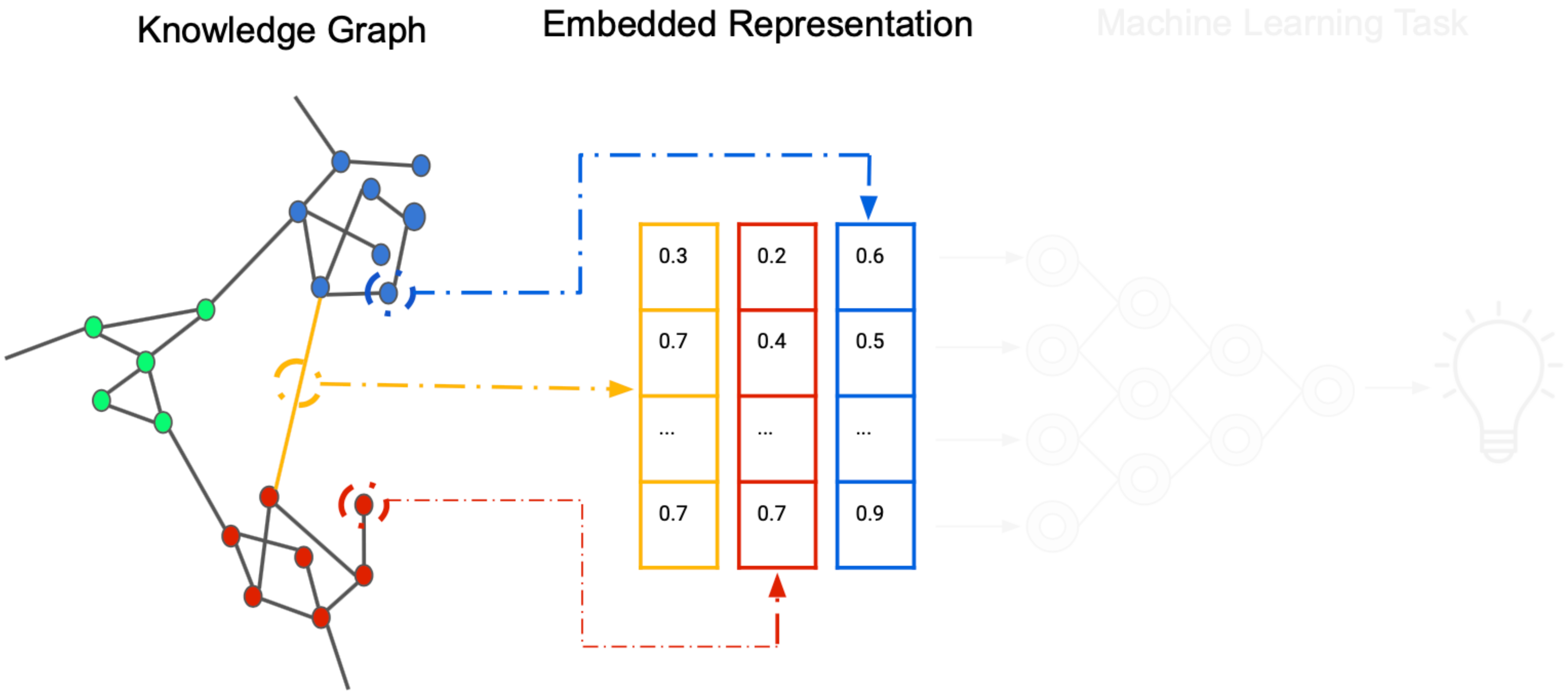


Image taken from Wikipedia: https://upload.wikimedia.org/wikipedia/commons/3/3f/KnowledgeGraphEmbedding.png

Introduction

Why do we need Entity Embeddings?

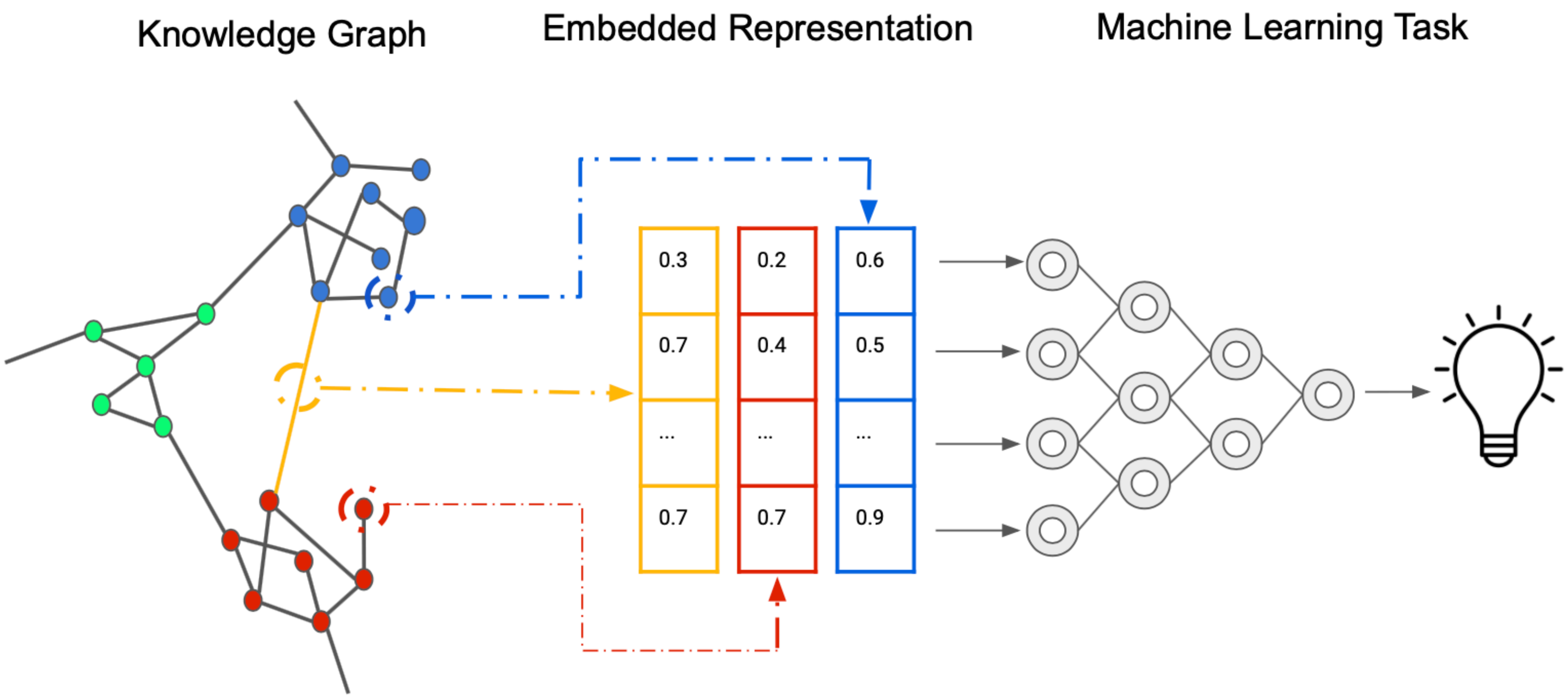


Image taken from Wikipedia: https://upload.wikimedia.org/wikipedia/commons/3/3f/KnowledgeGraphEmbedding.png

Introduction

Research question

What is the effect of semantically enriching an IoT KG, based on the quality of entity embeddings learned from it?

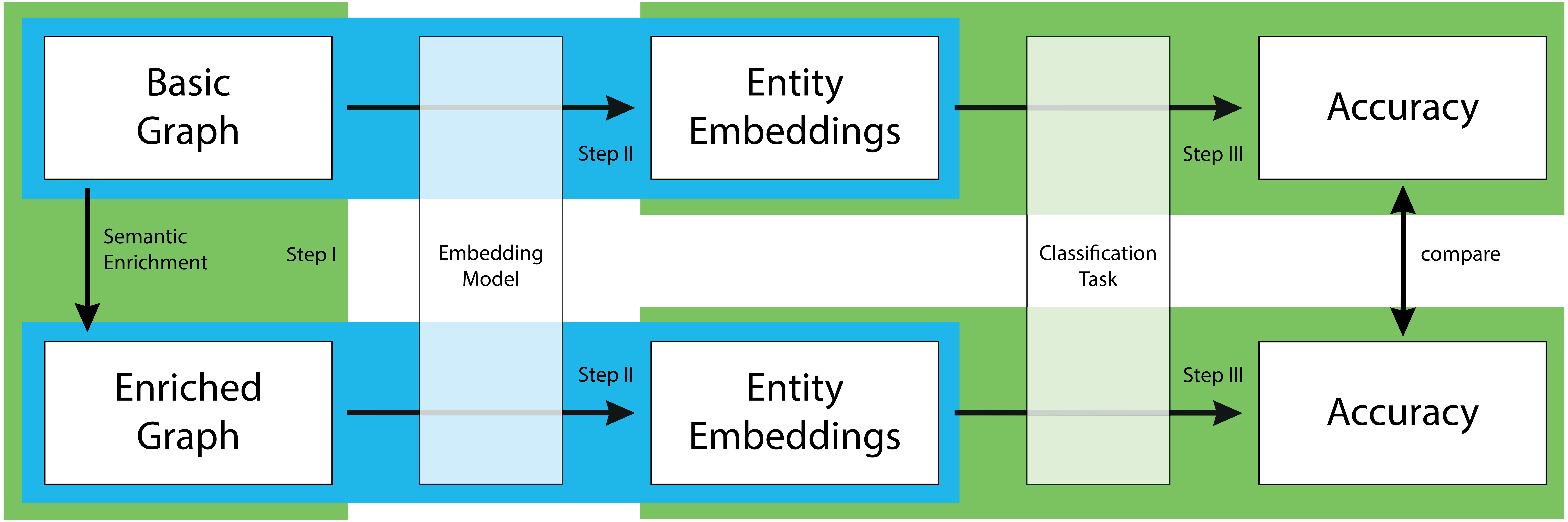
Introduction

Research question

What is the effect of semantically enriching an IoT KG, based on the quality of entity embeddings learned from it?

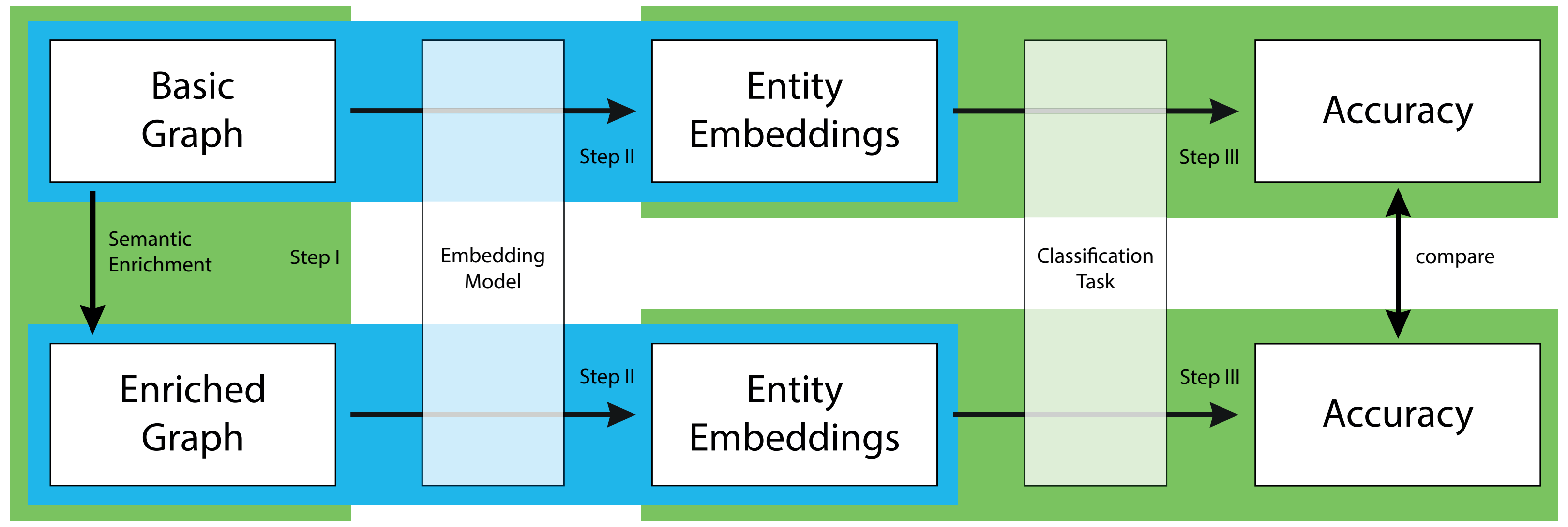
Experimental Pipeline

Overview



Experimental Pipeline

Overview



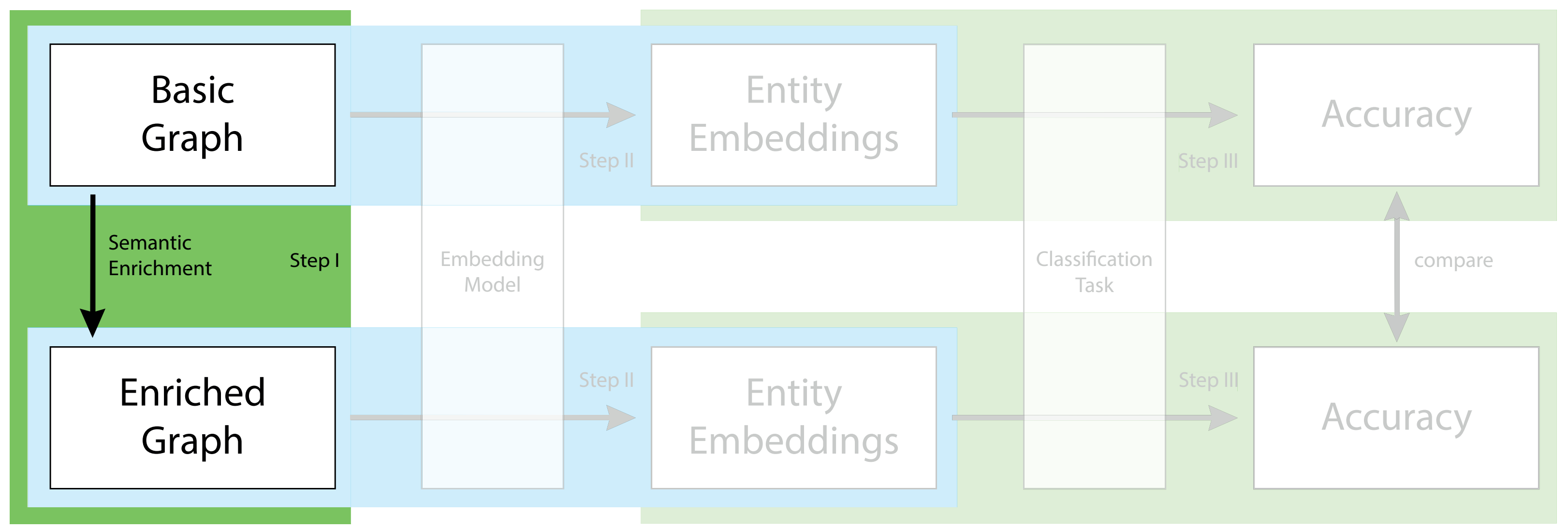
- Step 1: Semantic Enrichment

- Step 2: Embedding

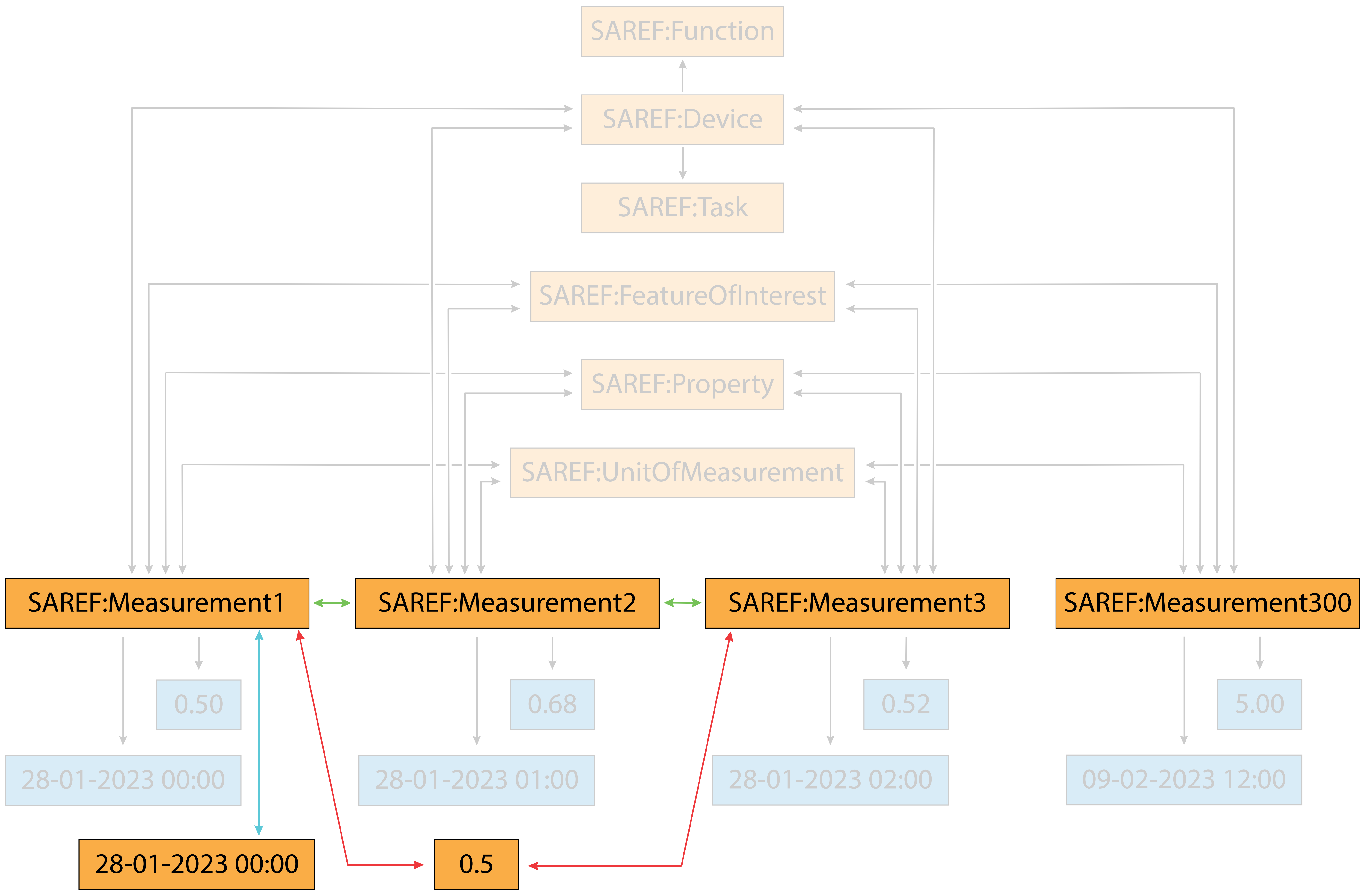
- Step 3: Evaluation

Experimental Pipeline

Semantic Enrichment step

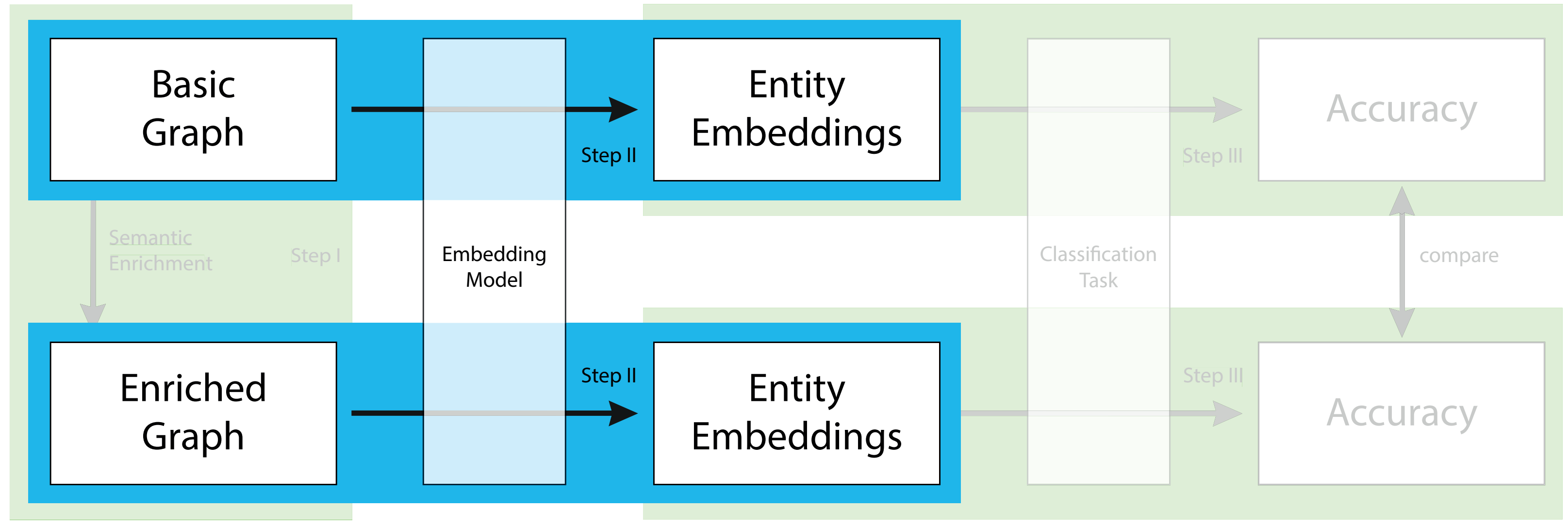


- Rounded value
- Sequence links
- Timestamp

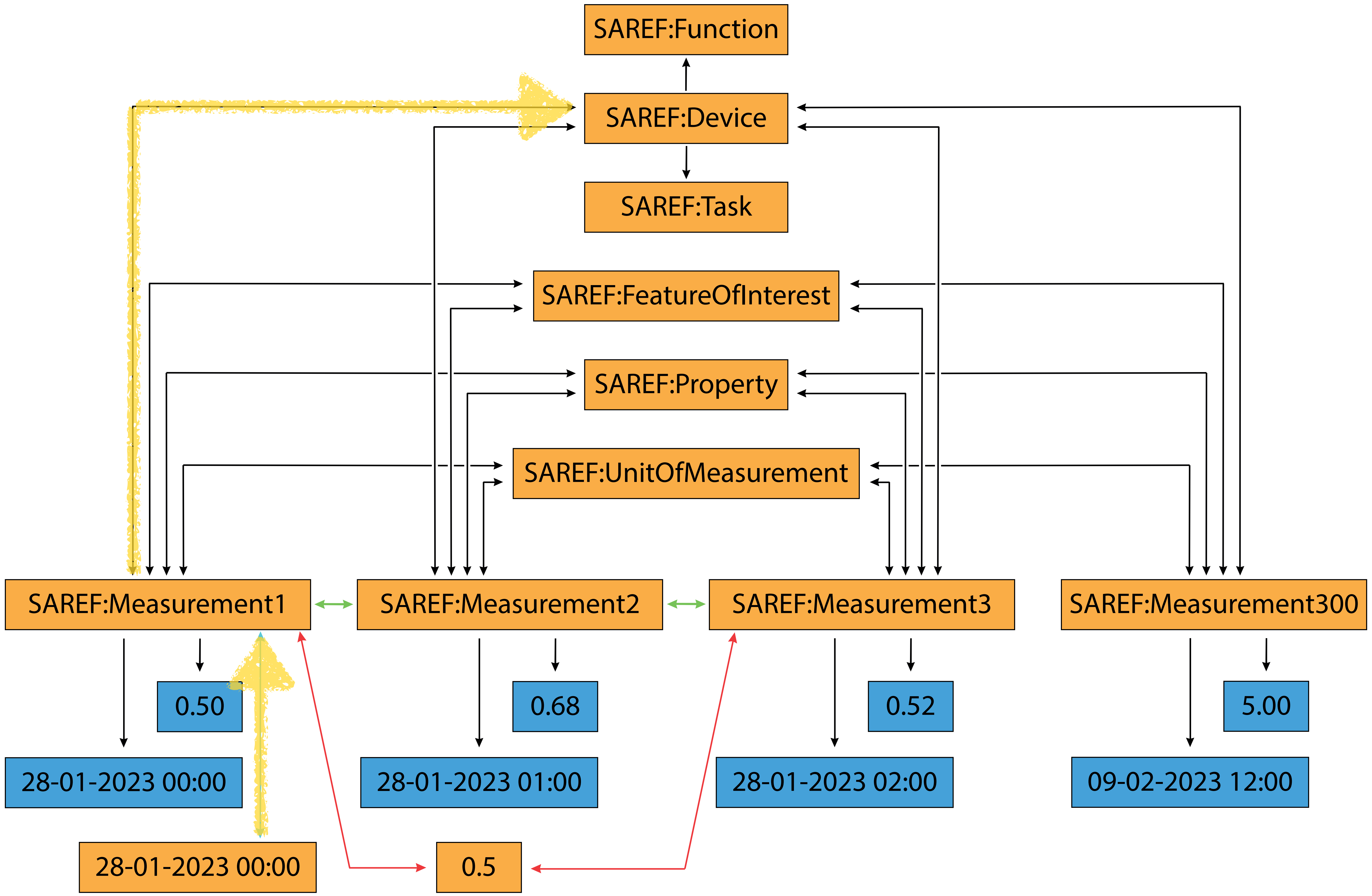


Experimental Pipeline

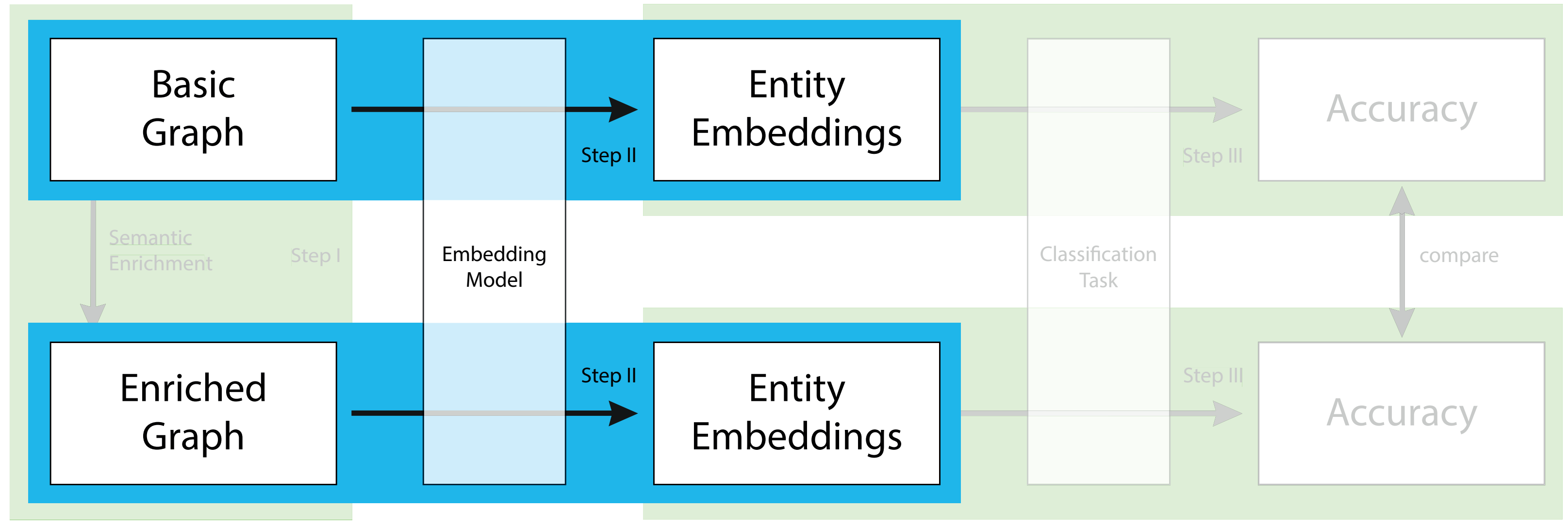
Embedding step



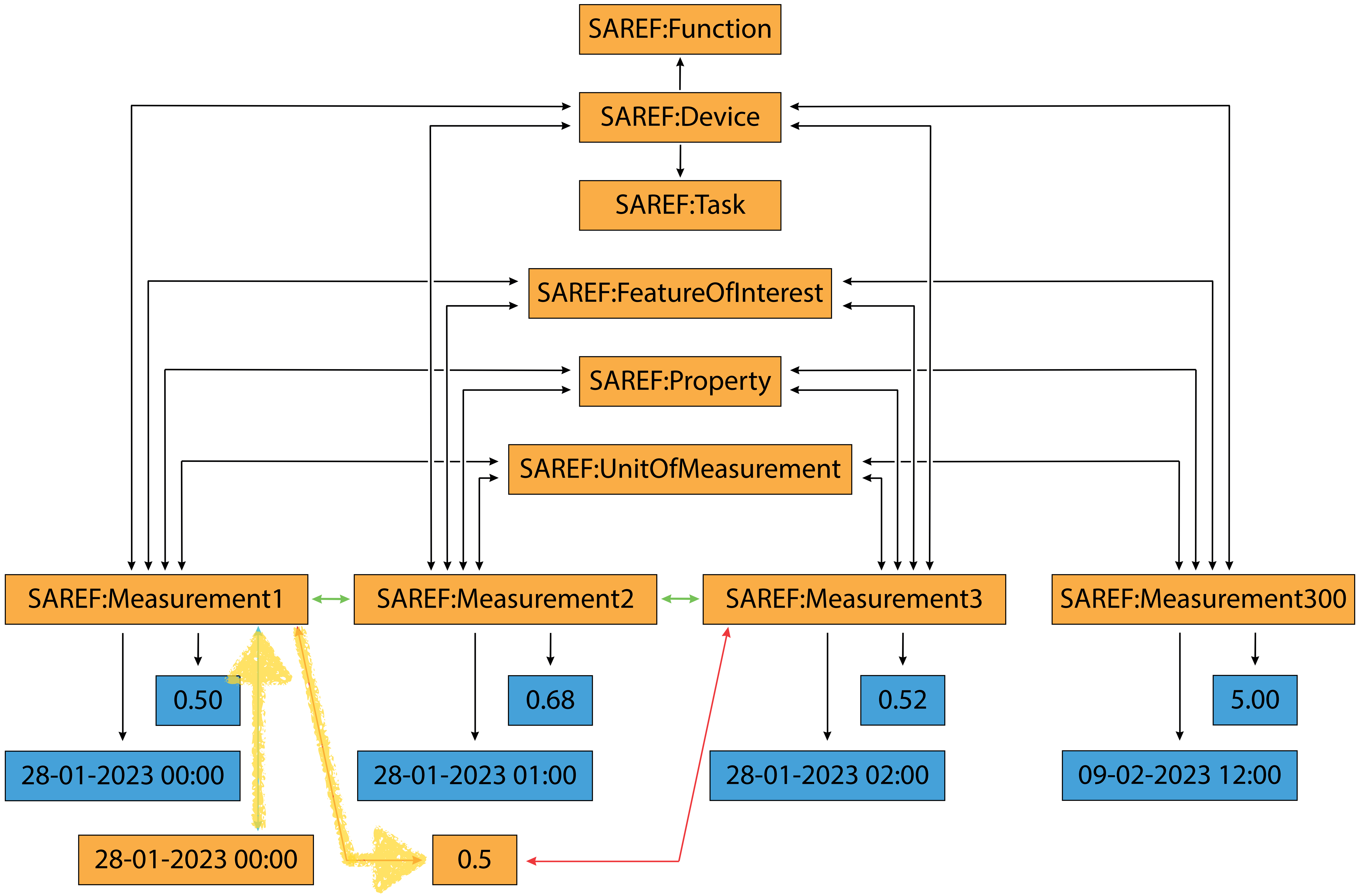
- RDF2vec
- Walk length of 2
- 25 walks per entity



Experimental Pipeline
Embedding step

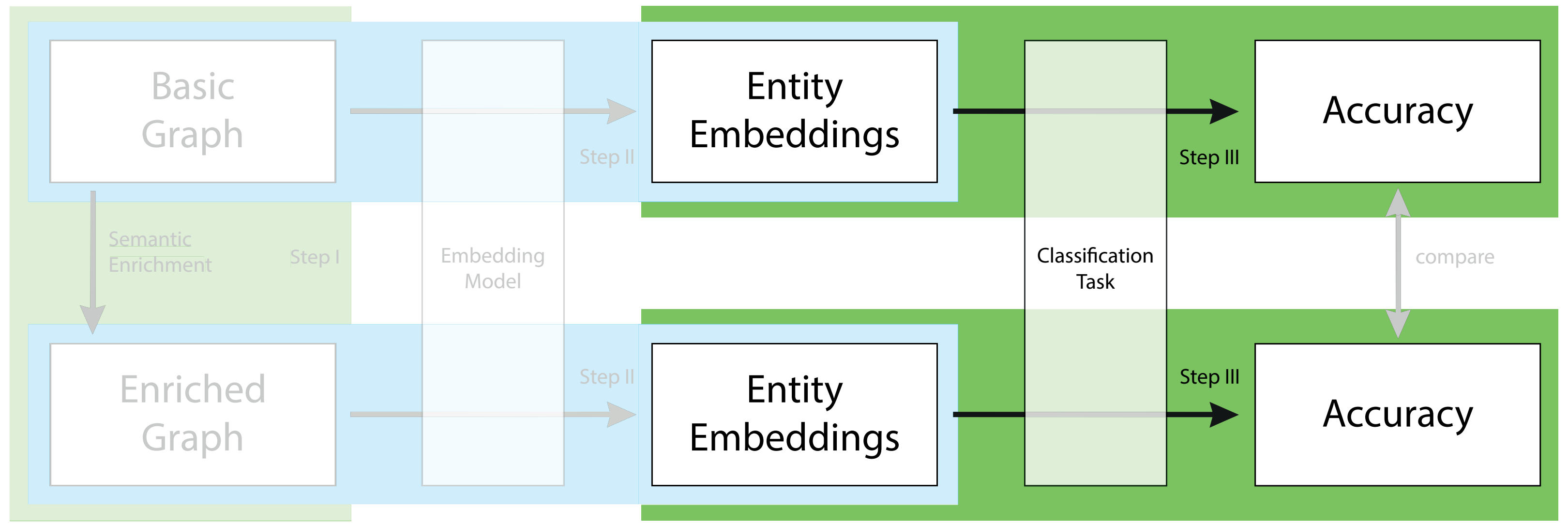


- RDF2vec
- Walk length of 2
- 25 walks per entity

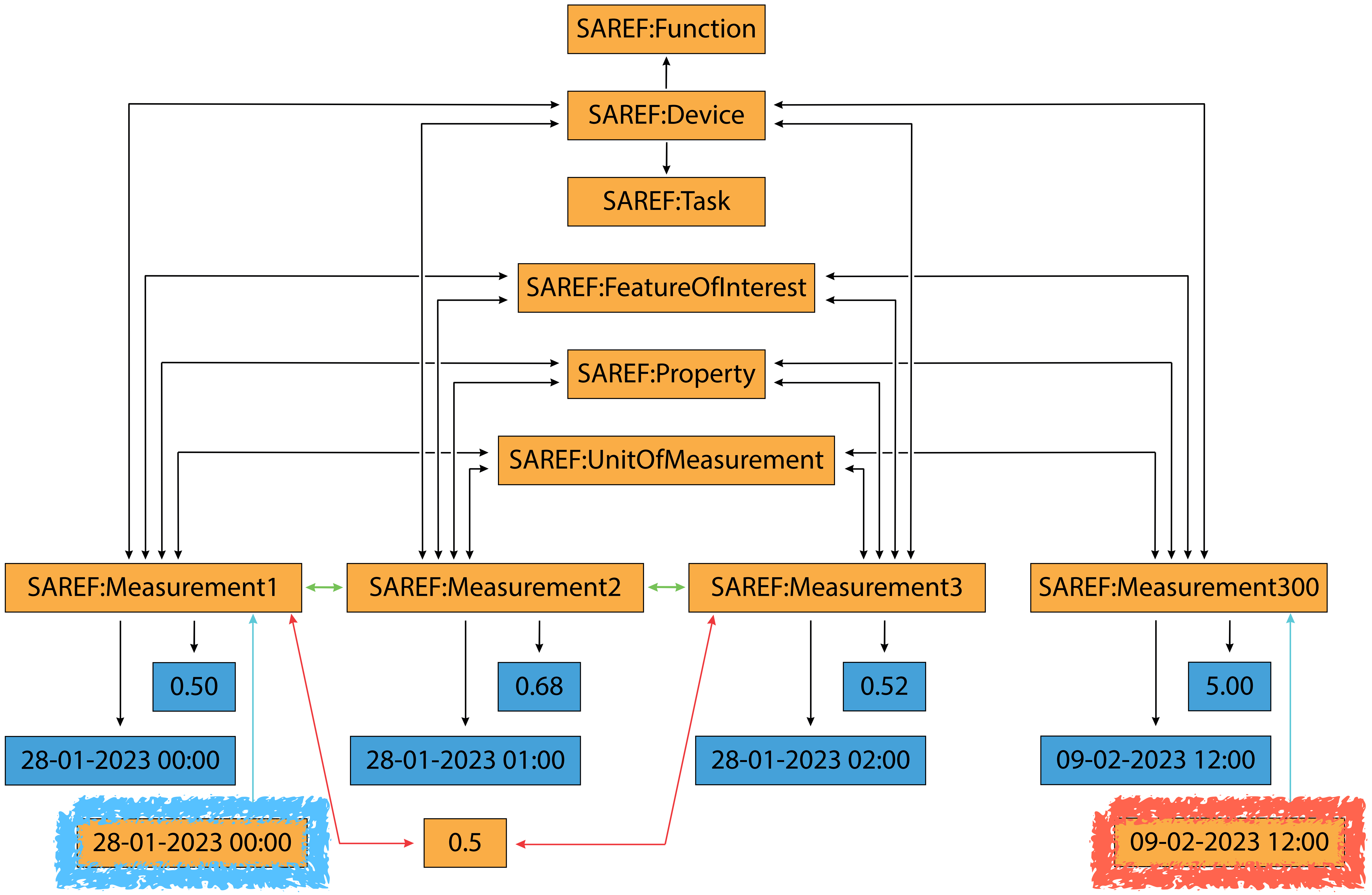


Experimental Pipeline

Evaluation step - Classification Task

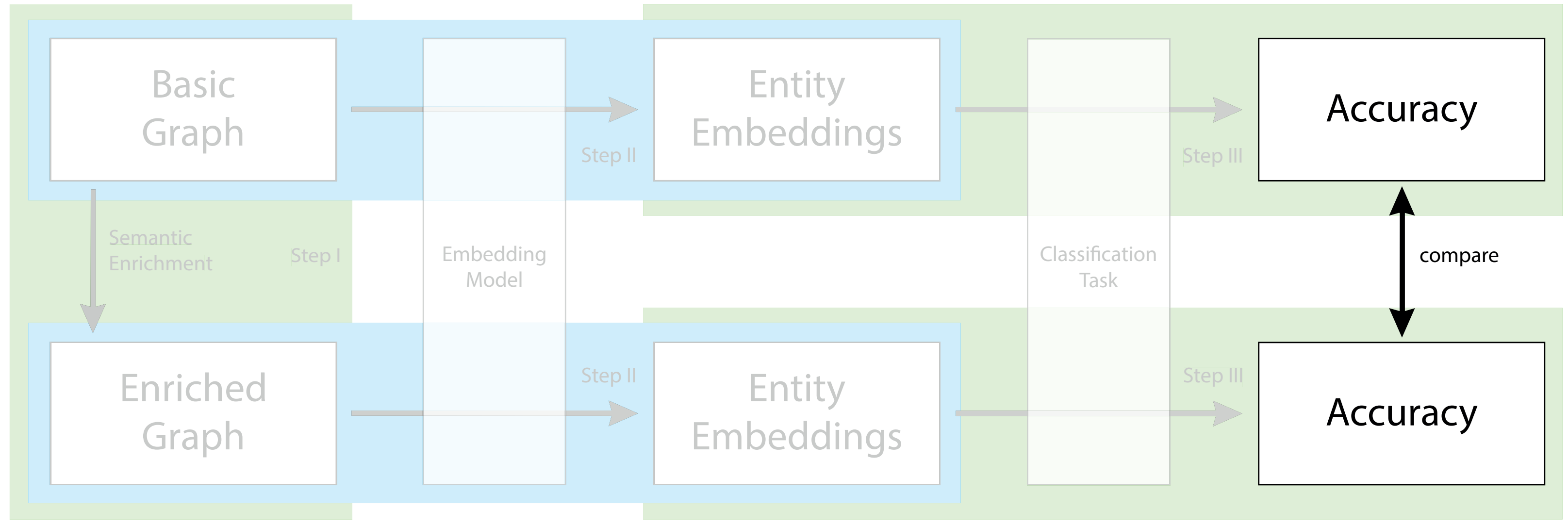


- Separated all timepoints based on outside temperature
- Labeled warmest half “warm” & coldest half “cold”
- Trained a MLP to classify timepoints based on label



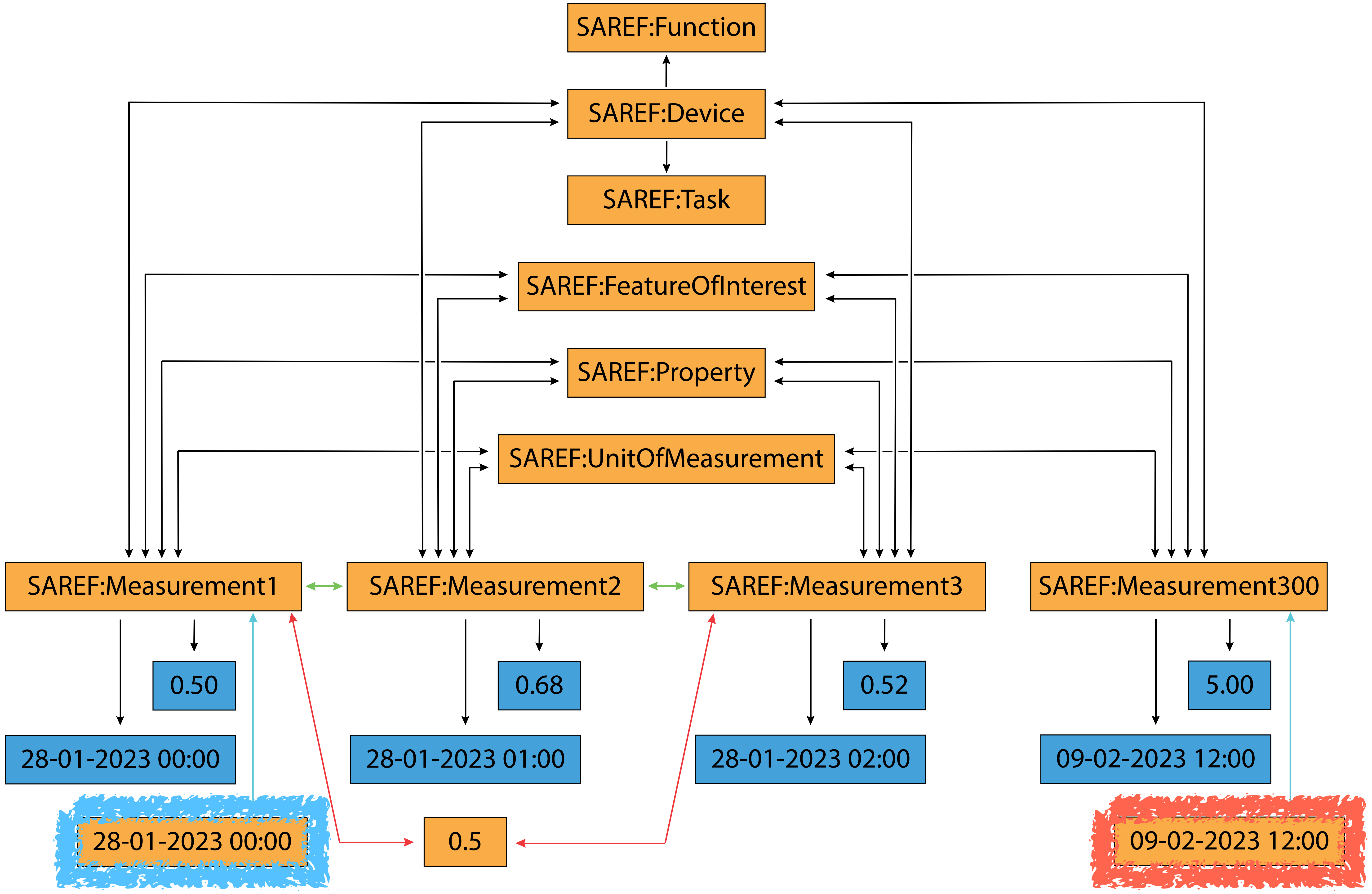
Experimental Pipeline

Evaluation step - Accuracy comparison

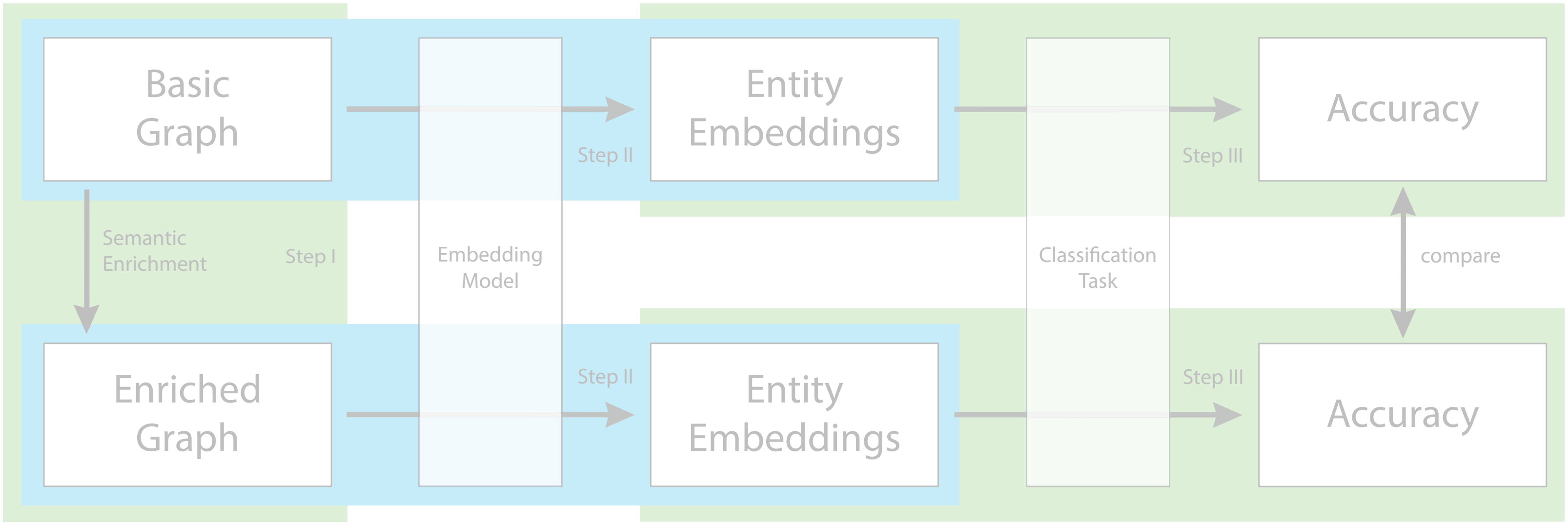


- Experiment is performed with both the original graph and enriched graph

- Accuracies are compared



Experimental Pipeline
Dataset

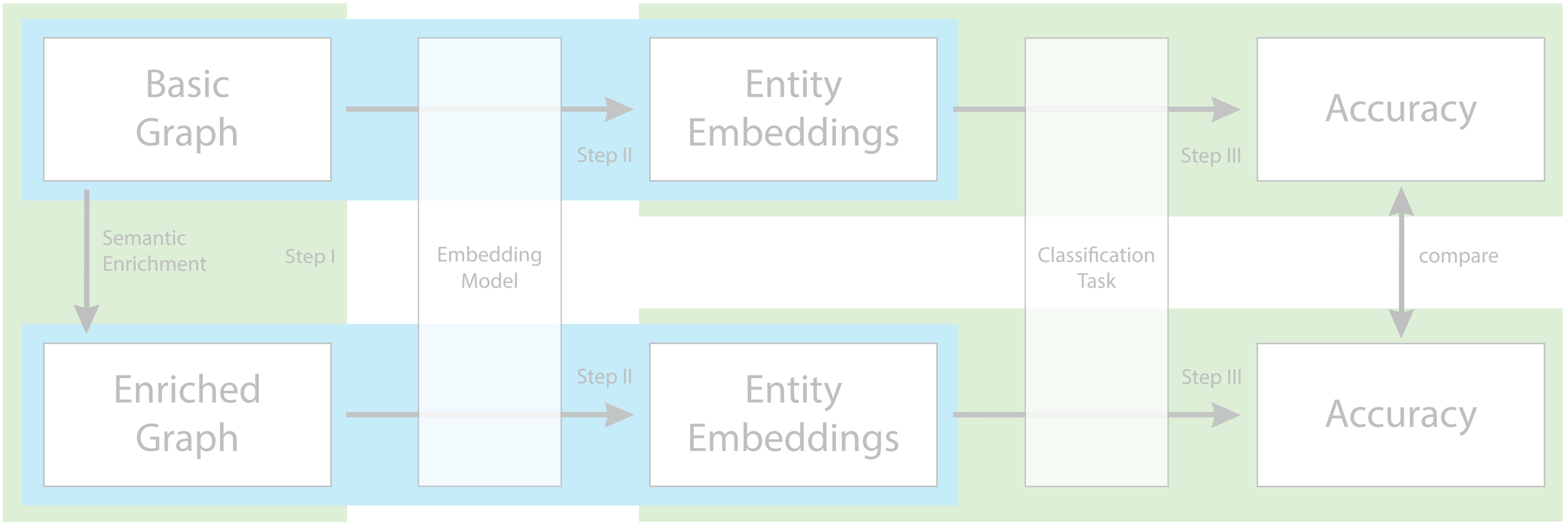


- OPSD Household dataset
- Hourly measurements
- 8133 timepoints (±11 months)
- Made into a IoT KG

	device types										
residence id	grid import	grid export	pv	dishwasher	ev	refridgerator	freezer	heat pump	washing machine	circulation pump	total
res1	1		1	1			1	1	1		6
res2	1			1			1		1	1	5
res3	1	1	1	1		1	1		1	1	8
res4	1	1	1	1	1	1	1	1	1		9
res5	1			1		1			1		4
res6	1	1	1	1			1		1	1	7
total	6	3	4	6	1	3	5	2	6	3	39

https://data.open-power-system-data.org/household_data/2020-04-15

Experimental Pipeline
Dataset



- **res1dev1:**
  1 device from 1 residence
- **res1devA**
  All devices from 1 residence
- **resAdevA**
  All devices from All residences

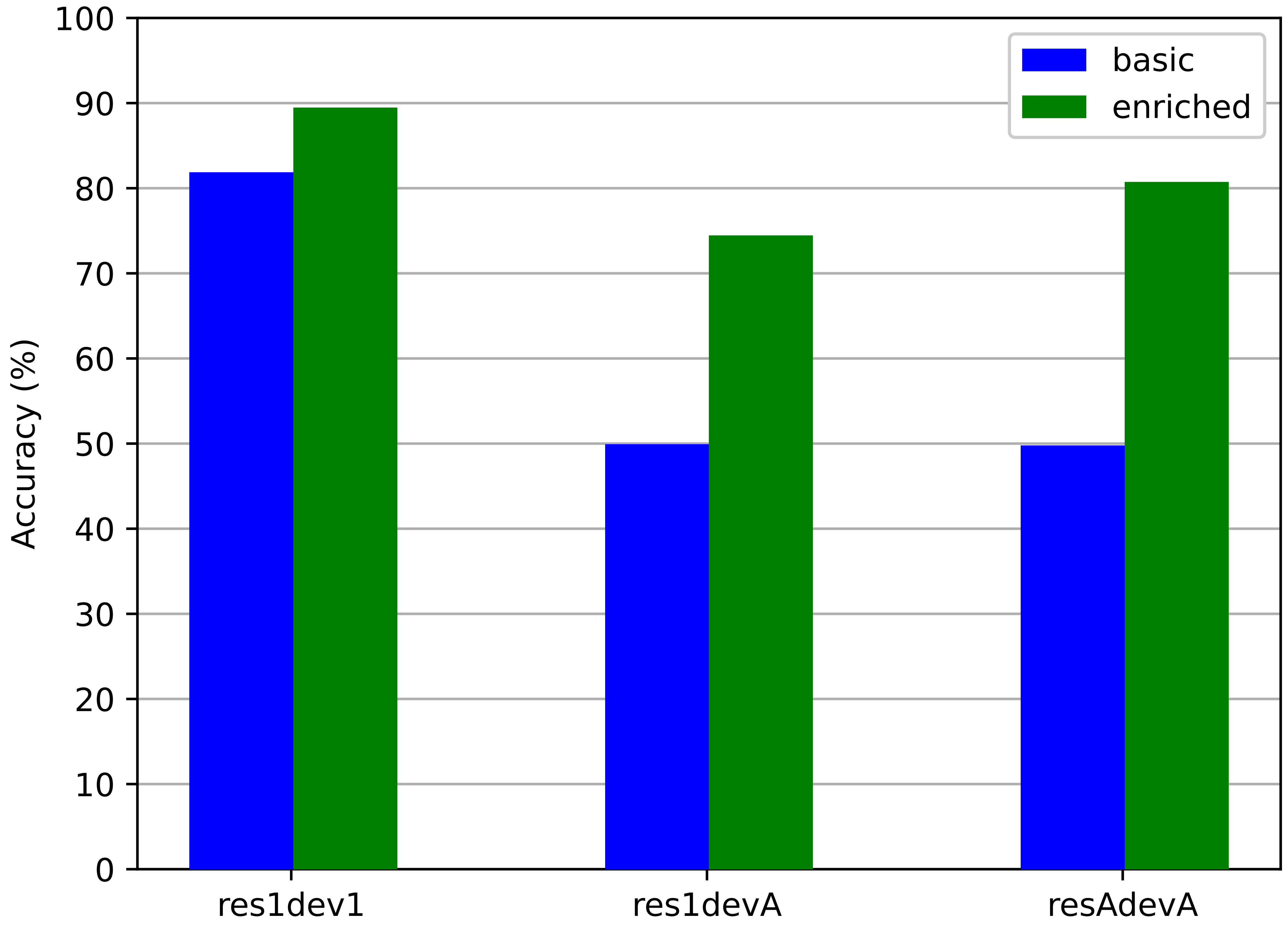
	device types										
residence id	grid import	grid export	pv	dishwasher	ev	refridgerator	freezer	heat pump	washing machine	circulation pump	total
res1	1		1	1			1	1	1		6
res2	1			1			1		1	1	5
res3	1	1	1	1		1	1		1	1	8
res4	1	1	1	1	1	1	1	1	1		9
res5	1			1		1			1		4
res6	1	1	1	1			1		1	1	7
total	6	3	4	6	1	3	5	2	6	3	39

https://data.open-power-system-data.org/household_data/2020-04-15

Results

Average accuracies of the classifiers

- Enriched outperforms basic every time
- More “useless” devices have a negative impact
- More “useful” devices have a positive impact



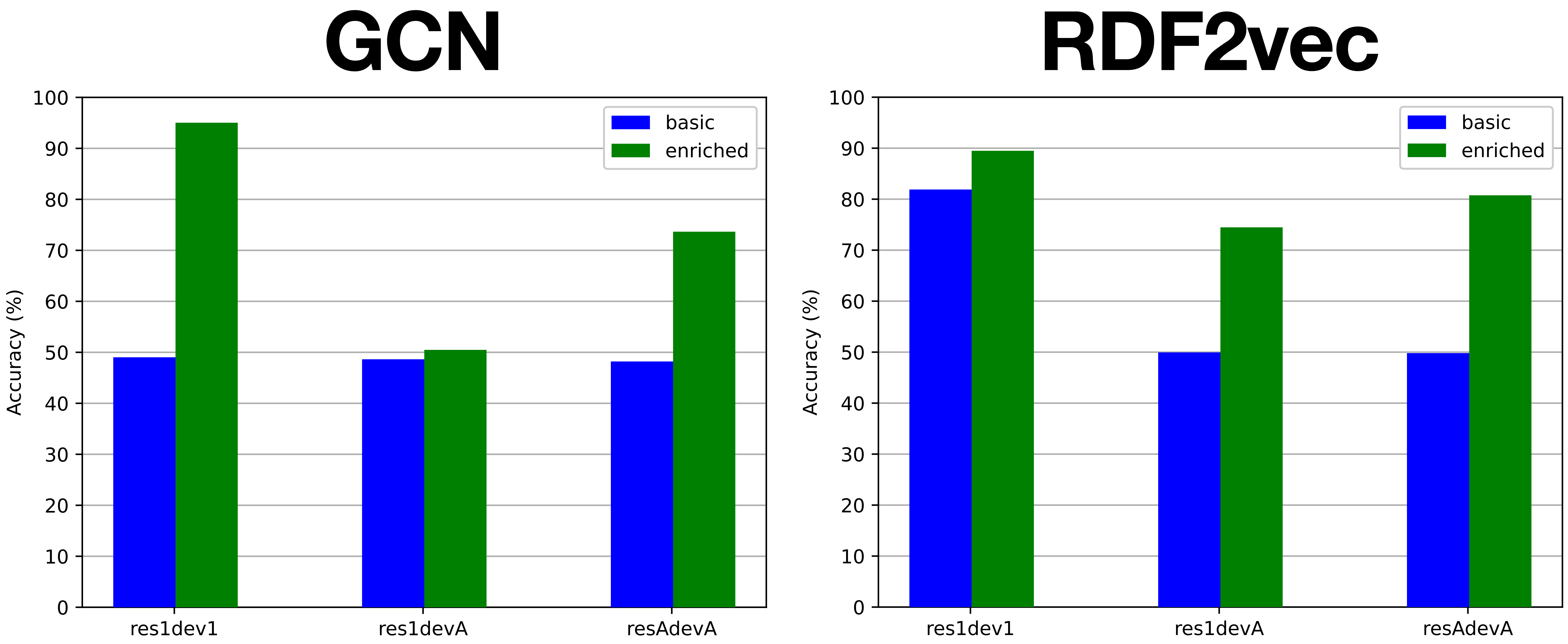
Discussion

Ongoing Research

- Using a different evaluation task
    - Value prediction
- Using a different embedding method
    - GCN
- Using different datasets
    - Pecan street (American consumption data)

Ongoing Research

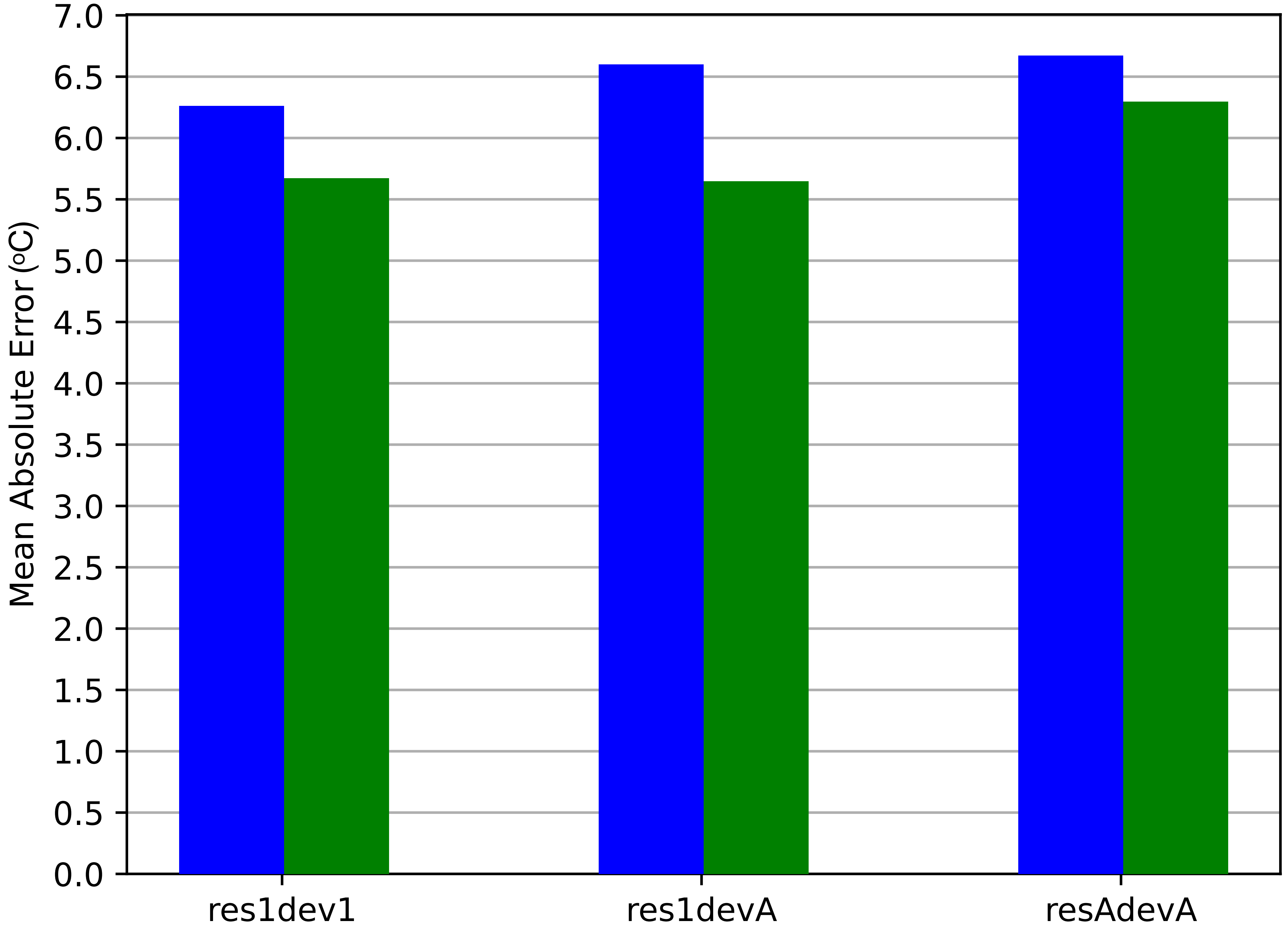
Average accuracies of the classifiers - RDF2vec & GCN



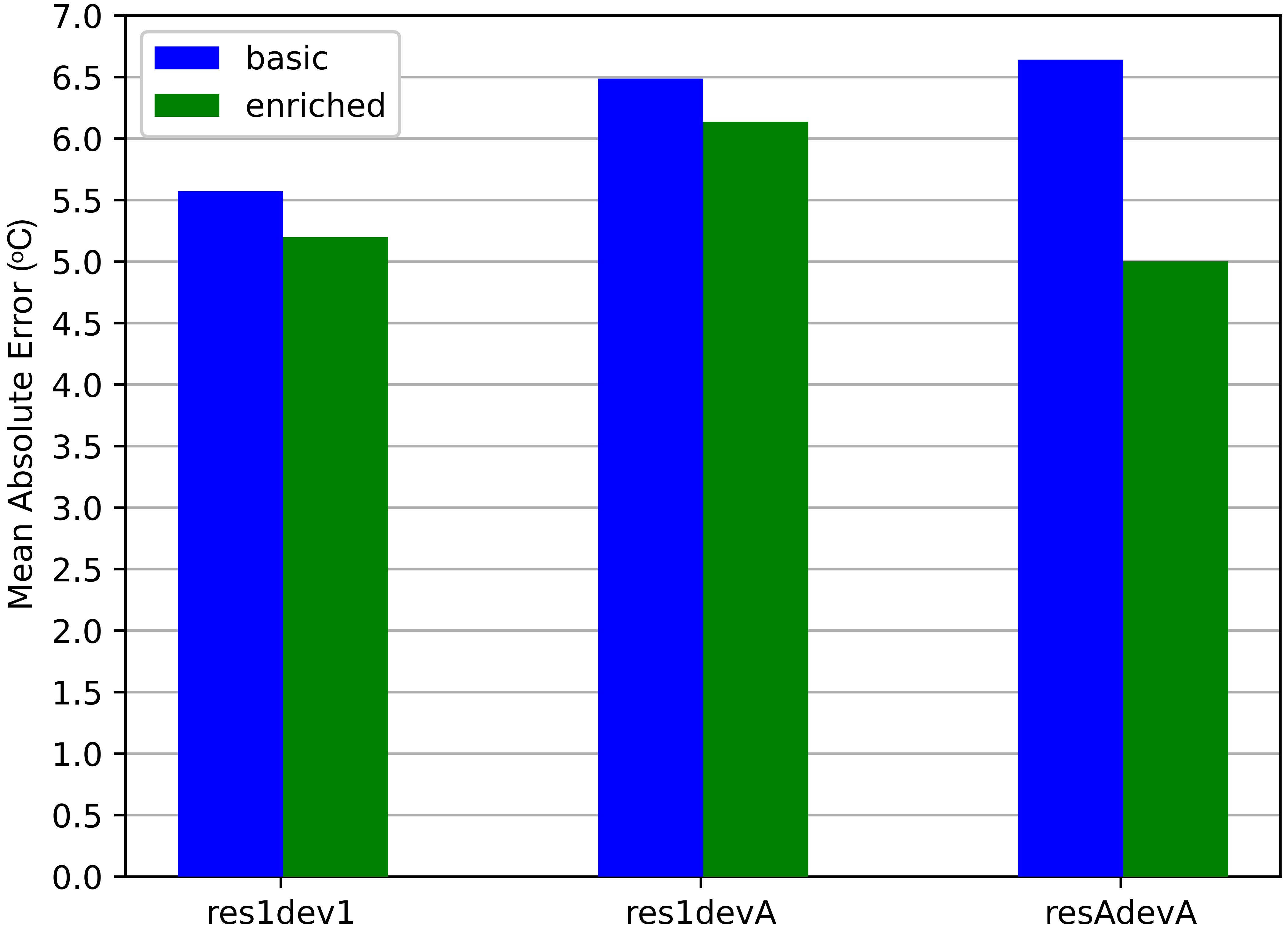
Ongoing Research

Average accuracies of the VALUE PREDICTOR - RDF2vec & GCN

GCN



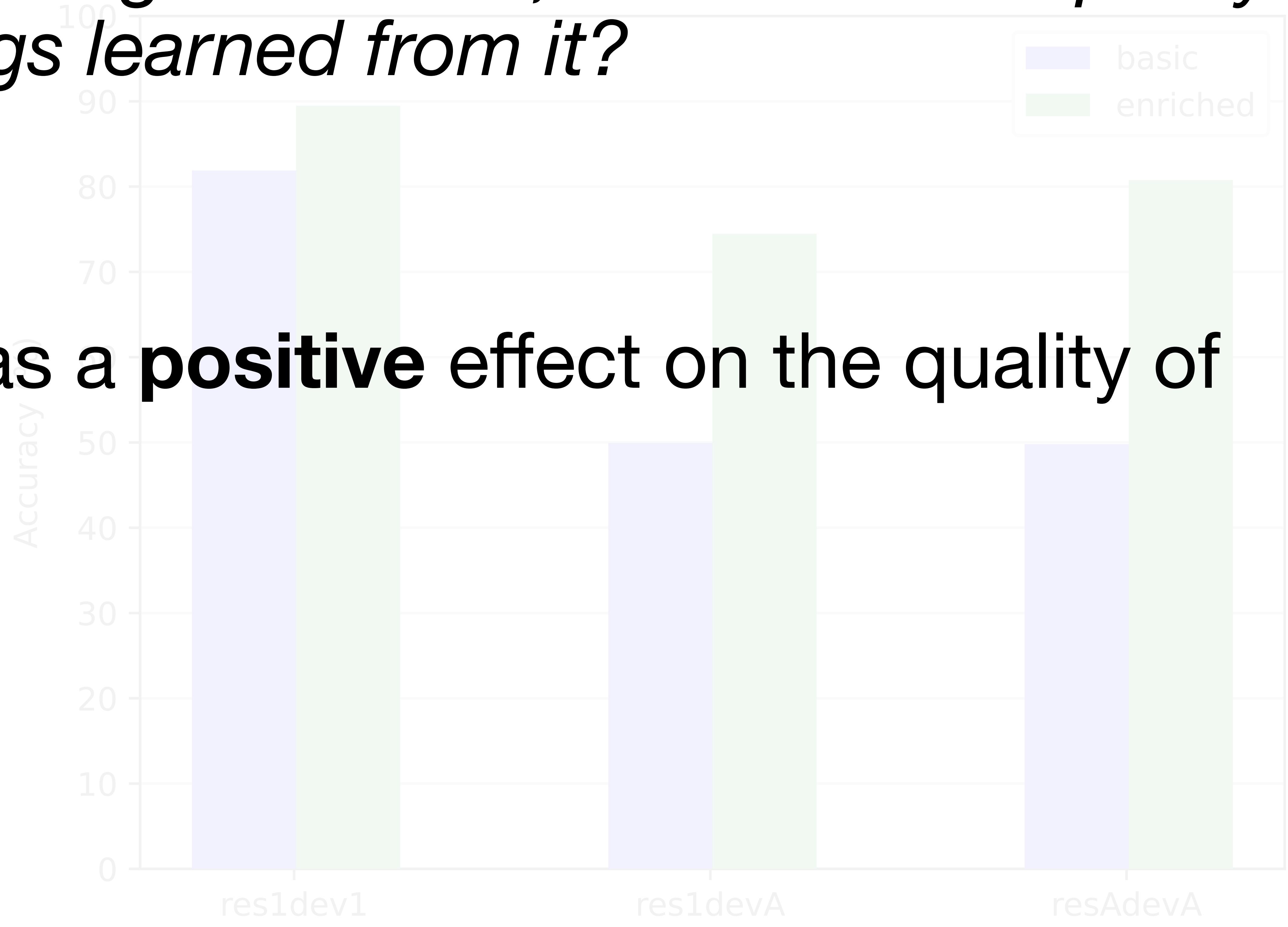
RDF2vec



Conclusion

What is the effect of semantically enriching an IoT KG, based on the quality of entity embeddings learned from it?

- In this setting, enriching an IoT KG has a **positive** effect on the quality of entity embeddings learned from it



Conclusion

What is the effect of semantically enriching an IoT KG, based on the quality of entity embeddings learned from it?

- In this setting, enriching an IoT KG has a **positive** effect on the quality of entity embeddings learned from it

Thank you for listening