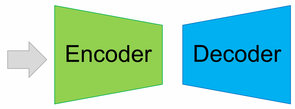
Automatic Comment Generation via Multi-Pass Deliberation
利用多轮改写机制的代码注释生成方法

沐方文，陈啸，石琳，王松，王青

The 37th IEEE/ACM International Conference on Automated Software Engineering

Background

```
public CategoricalTable copy(){
  Map<Value,Double> newTable=new HashMap<Value,Double>();
  if(variable == null){
    variable = 1;
  }
  if(table.isEmpty()){
    return new CategoricalTable(variable);
  }
  for(Value v : table.keySet()){
    newTable.put(v,table.get(v));
  }
  return new CategoricalTable(variable,newTable);
}
```

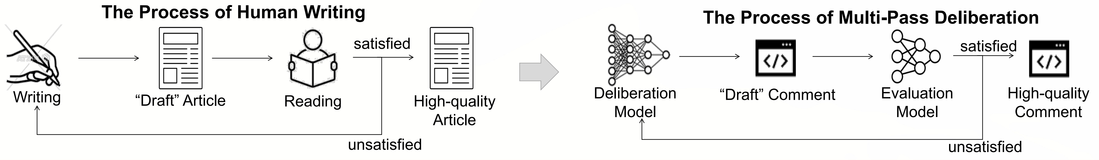


Ground Truth:
returns a copy of the probability table
Baseline 1:
returns a copy of the bytes written to this stream
Baseline 2:
creates a new copy of the given table of the table

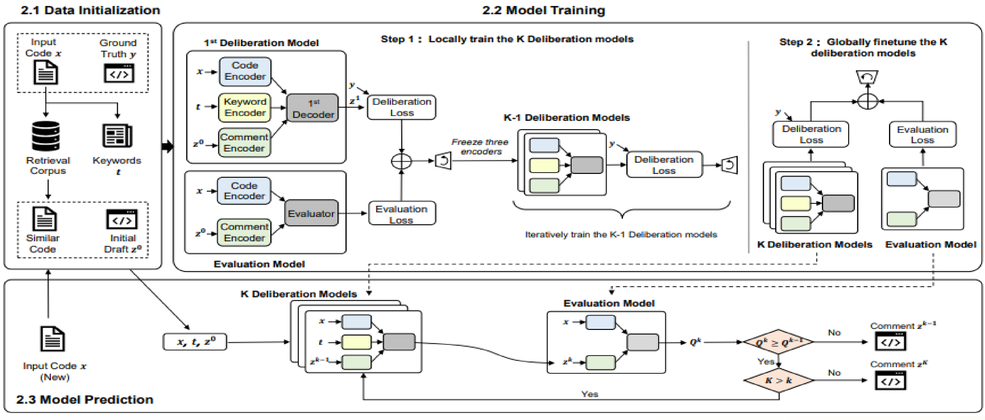
Limitations of the existing approaches:

- *Cannot correct the mispredicted words*
- *Cannot leverage the global information*

Motivation



Approach: **DECOM**



Three Stages:

- **Stage1: Data Initialization**, for extracting identifier names from code and retrieving the initial draft.
- **Stage2: Model Training**, for leveraging a two-step training strategy to optimize DECOM.
- **Stage3: Model prediction**, for generating the target comment of the new source code.

Evaluation

Method	JCSD							PCSD						
	BLEU-1/2/3/4				ROUGE-L	METEOR	CIDEr	BLEU-1/2/3/4				ROUGE-L	METEOR	CIDEr
LSI	31.4	22.5	19.3	17.3	34.8	14.4	1.803	36.3	23.6	20.1	17.6	40.0	17.2	1.982
VSM	33.3	24.4	21.1	19.0	36.6	15.4	1.983	38.9	26.1	22.1	19.3	42.7	19.0	2.216
NNGen	33.0	24.4	20.9	18.7	36.3	15.0	1.933	36.5	23.8	20.1	17.4	40.2	17.1	1.967
CODE-NN	23.9	12.8	8.6	6.3	28.9	9.1	0.978	30.8	15.4	10.7	8.1	35.1	13.4	1.229
TL-CodeSum	29.9	21.3	18.1	16.1	33.2	13.7	1.660	31.1	16.5	12.5	10.4	35.3	13.6	1.335
Hybrid-DRL	32.4	22.6	16.3	13.3	26.5	13.5	1.656	41.1	26.2	19.5	15.0	42.2	17.9	2.042
Re^{2}com	33.7	23.6	19.0	16.3	38.1	15.1	1.807	36.6	22.3	17.4	14.5	40.8	17.0	1.813
Rencos	37.5	27.9	23.4	20.6	42.0	17.3	2.209	43.1	29.5	24.2	20.7	47.5	21.1	2.449
EditSum	34.1	24.3	19.5	16.9	38.6	15.2	1.865	37.7	23.1	18.2	15.6	42.0	17.1	1.894
DECOM	**40.4**	**30.2**	**25.2**	**22.3**	**44.5**	**19.6**	**2.442**	**45.6**	**31.4**	**25.5**	**21.9**	**49.3**	**22.5**	**2.603**

RQ1: Comparison with Baselines

- Compared to the best baseline Rencos, DECOM improves the performance of BLEU-4, ROUGE-L, METEOR, and CIDEr by 8.3%, 6.0%, 13.3%, and 10.5% on JCSD dataset, by 5.8%, 3.8%, 6.6%, and 6.3% on PCSD dataset, respectively.

Variants	JCSD							PCSD						
	BLEU-1/2/3/4				ROUGE-L	METEOR	CIDEr	BLEU-1/2/3/4				ROUGE-L	METEOR	CIDEr
DECOM w/o Multi-pass Deliberation	38.9	28.5	23.5	20.8	43.1	18.8	2.274	43.5	29.3	23.8	20.4	47.5	21.1	2.424
DECOM w/o Evaluation Model	39.5	29.3	24.3	21.5	43.7	19.0	2.338	44.6	30.3	24.3	20.6	48.6	21.6	2.478
DECOM	**40.4**	**30.2**	**25.2**	**22.3**	**44.5**	**19.6**	**2.442**	**45.6**	**31.4**	**25.5**	**21.9**	**49.3**	**22.5**	**2.603**

RQ2: Component Analysis

- Both the multi-pass deliberation and evaluation model components have positive contributions to the performance of DECOM, where the multi-pass deliberation component contributes more to increasing the performance.

Conclusion

- We propose a **multi-pass deliberation framework** for comment generation, named DECOM, which is **inspired by the human cognitive process**, and can effectively generate comments in an iterative way.
- We conduct an experimental evaluation of the performance of DECOM against SOTA baselines, which shows that **DECOM outperforms all baselines by a large margin** in main evaluation metrics.

https://github.com/ase-decom/ASE22_DECOM