

Developer-Intent Driven Code Comment Generation

Fangwen Mu^{*†‡}, Xiao Chen^{*†‡}, Lin Shi^{*†‡§}, Song Wang[¶], Qing Wang^{*†‡}

^{*} State Key Laboratory of Intelligent Game, Beijing, China

[†] Science and Technology on Integrated Information System Laboratory,
Institute of Software Chinese Academy of Sciences, Beijing, China

[‡] University of Chinese Academy of Sciences, Beijing, China

[¶] Lassonde School of Engineering, York University, Toronto, Canada

{fangwen2020, chenxiao2021, shilin, wq}@iscas.ac.cn, wangsong@yorku.ca

発表者中谷祐太

開発者のコメント意図

コードスニペット

```
public synchronized boolean start() {  
    if (!isStarted()) {  
        final ExecutorService tempExec;  
        executor = getExternalExecutor();  
        if (executor == null) {  
            executor = tempExec = createExecutor();  
        } else {  
            tempExec = null;}  
        future = executor.submit(createTask(tempExec));  
        return true;}  
    return false;  
}
```

開発者が書いたコメント

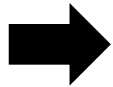
What: Starts the background initialization.

Why: With this method the initializer becomes active and invokes the initialize() method in a background task.

How-it-is-done: Get an external executor to create a background task. If there is not any, it creates a new one.

How-to-use: After the construction of a BackgroundInitializer() object it start() method has to be called.

Property: Return a flag whether the initializer could be started successfully.

- 
- 5つの文から構成
 - 5つの異なる意図が存在

目視調査より

- 1つのコメントに平均2.81の異なる意図が存在

コメント意図の分類

カテゴリー	説明	例
What	メソッドの機能	“A helper function that process the stack.”
Why	メソッドを提供する理由,またはメソッドの設計根拠	“Get a copy of the map (for diagnostics)”
How-to-use	メソッドの使用方法,またはメソッドを使用する際に予想されるセットアップ	“Should be called before the object is used”
How-it-is-done	メソッドの実装の詳細	“Convert the byte[] to a secret key”
Property	メソッドの事前条件や事後条件など,メソッドのプロパティ	“Wait until seqno is greater than or equal to the desired value or we exceed the timeout.”
Others	不特定または曖昧なコメント	“The implementation is awesome.”

既存手法の課題

コードスニペット

```
public synchronized boolean start() {  
    if (!isStarted()) {  
        final ExecutorService tempExec;  
        executor = getExternalExecutor();  
        if (executor == null) {  
            executor = tempExec = createExecutor();  
        } else {  
            tempExec = null;}  
        future = executor.submit(createTask(tempExec));  
        return true;}  
    return false;  
}
```

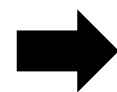
最先端のモデルが生成したコメント

What the SOTA methods generate:

Rencos (Zhang et al. ICSE20): starts the execution of the executor

Editsum (Li et al. ASE21): starts the thread

AST-Trans (Tang et al. ICSE22): starts the task



- 1つの文で構成
- メソッドの機能のみ記述

課題

- 開発者の意図を考慮できず実用的なニーズを満たせない

目的・アプローチ

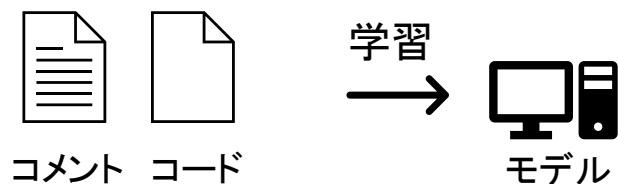
目的

様々な意図を満たすコメントを生成するアプローチを開発する

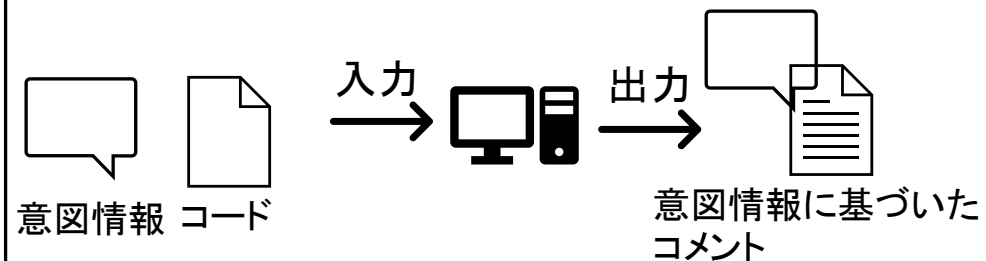
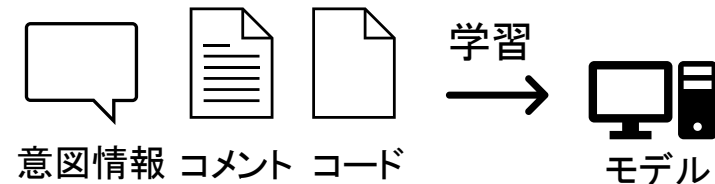
アプローチ

意図情報を考慮したモデルDOMEを提案

既存モデル

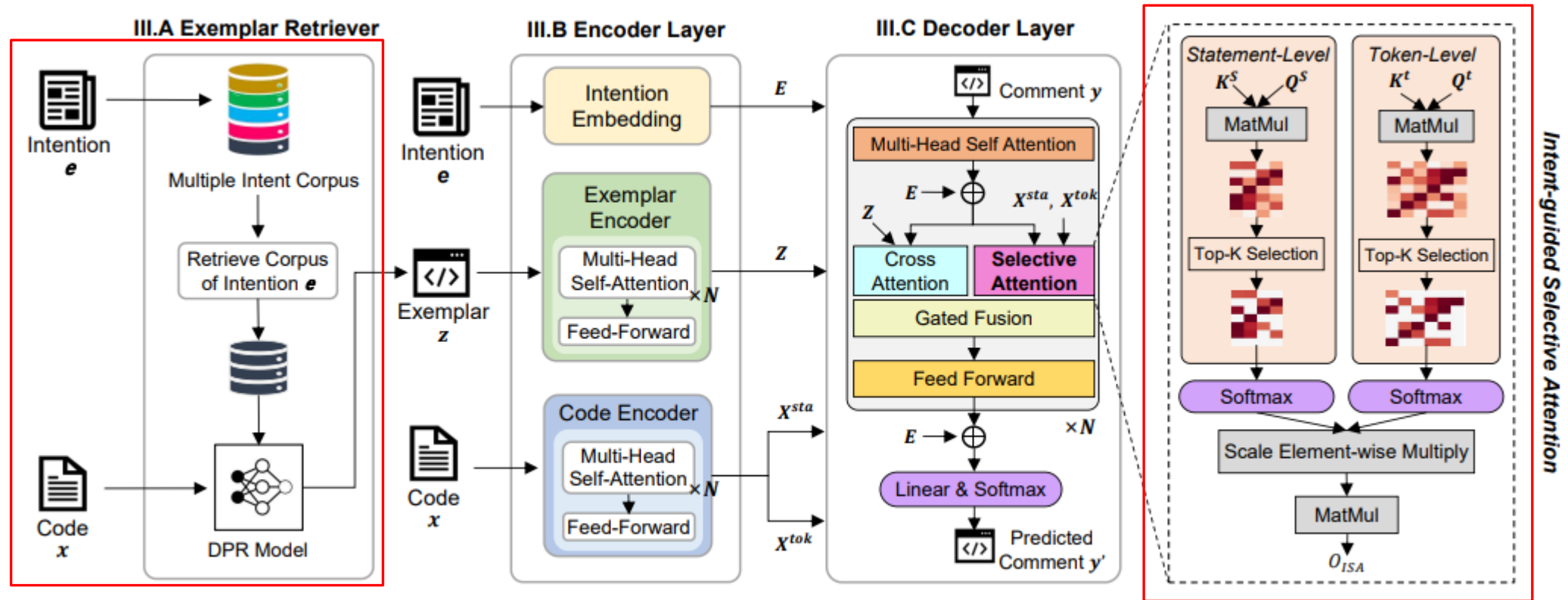


DOME



DOME

(Developer-intent driven cOde coMment gEneration)



DRPモデルで類似度が
高いコメントを模範解答とする

Fig. 2: The architecture of DOME

貢献度の高い情報に注意を集中させる

COIN(Comment-INtent)

コメント意図の自動分類器

作成方法

1. データセット作成

- FuncomとTLCから10kずつコードとコメントのペアをランダムにサンプリング
- 手動で意図カテゴリを分類

2. 上記データセットを用いCodeBERTをファインチューニング

TABLE II: Statistics of the manually-labeled intent dataset

Category	Count	Proportion
What	12,264	61.32%
Why	1,708	8.54%
How-to-use	573	2.87%
How-it-is-done	2,933	14.67%
Property	2,270	11.35%
Others	252	1.26%

TABLE III: The performance of intent classification for code comments

Method	What			Why			How-to-use			How-it-is-done			Property			Others			Macro-Average		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
LGBM	87.4	92.3	89.8	81.5	76.8	79.0	89.9	72.5	80.1	71.6	65.8	68.5	90.8	89.7	90.2	12.6	6.4	8.4	72.3	67.2	69.3
RF	83.9	94.3	88.8	84.0	73.2	78.2	92.3	70.3	79.6	75.8	54.5	63.4	90.9	89.3	90.0	39.9	29.7	30.3	77.8	68.5	71.7
DT	86.3	87.0	86.6	73.7	74.0	73.8	74.0	69.5	71.5	63.0	59.9	61.3	87.8	86.8	87.2	16.6	32.5	20.6	68.3	66.9	66.8
CNN	87.4	93.3	90.2	82.7	77.3	79.9	90.4	71.3	79.5	71.8	65.7	68.6	94.5	89.4	91.8	82.1	29.2	41.5	84.8	71.0	75.3
BiLSTM	88.4	89.0	88.6	78.5	74.6	76.4	81.0	62.8	70.4	60.5	68.8	64.2	96.1	88.2	92.0	10.0	0.9	1.7	69.1	64.1	65.5
BiLSTM-Attn	87.3	91.5	89.3	79.5	75.2	77.2	80.2	68.1	73.3	68.2	67.3	67.7	96.1	88.6	92.1	0.0	0.0	0.0	68.5	65.1	66.6
COIN	92.5	93.5	93.0	87.1	86.1	86.6	93.0	86.6	89.6	78.6	79.2	78.9	96.4	92.9	94.6	99.3	90.7	94.7	91.1	88.2	89.6

データセット

使用データセット

- Funcom

29KのJavaプロジェクトからの210万のコード-コメントペア

- TLC

Java GitHubリポジトリから収集された87,136のコード・コメント・ペア

前処理

- 自動クリーニング・ツールCAT2によってクリーニング
- COINを意図ラベルを追加
- ラベルがothersのものを削除
- TLCのテストセットから重複するコードとコメントのペアを削除

TABLE IV: Statistic of Funcom and TLC datasets

Dataset	Funcom	TLC
Train	1,178,923	53,528
Valid	62,383	7,555
Test	69,259	4,985
What	762,884	36,604
Why	168,912	7,708
How-to-use	27,543	1,085
How-it-is-done	166,286	14,392
Property	184,940	6,279

RQ

RQ1: DOME の性能は,最新のコメント生成ベースラインと比較してどうか？

RQ2: DOME の各コンポーネントは,全体的な性能にどのように貢献しているか？

RQ3: DOME によって生成された意図認識コメントの知覚品質はどの程度か？

RQ1: ベースラインとの比較

実験方法

- データセットを各意図カテゴリごとに分け,すべてのベースラインとDOMEを同じ訓練セットで訓練し,カテゴリごとにテストし,カテゴリごとに評価する.また,全カテゴリの平均性能も計算する

Intent	Method		Funcom			TLC		
			BLEU	ROUGE-L	METEOR	BLEU	ROUGE-L	METEOR
Average	Baseline	Hybrid-DRL	21.33	27.68	11.09	16.49	30.00	11.88
		CodeTrans	25.25	35.05	15.78	18.30	33.71	14.42
		Re ² Com	23.04	30.58	13.09	19.17	32.37	13.81
		Rencos	24.39	30.77	14.22	20.17	33.01	14.33
		EditSum	25.80	31.02	14.18	18.59	30.80	12.35
		AST-Trans	25.33	36.41	17.30	19.42	29.74	14.61
	Our Approach	DOME	31.83	42.45	20.48	22.20	36.67	16.47
		DOME w/o ISA	30.52	40.09	19.24	21.47	35.61	15.18
		DOME w/o ER	30.09	39.57	18.92	20.58	34.59	14.44

結果

- DOMEは全てのカテゴリ,全てのデータセット,全ての評価指標で最良の結果を得た

RQ2 : コンポーネント分析

実験方法

- ISAの代わりに一般的なクロスアテンションを用いてコメントを生成, 評価
- 模範解答を削除し, エンコーダとデコーダのフレームワークのみを使ってコメントを生成, 評価
- RQ1でのDOMEの結果と比較

Intent	Method		Funcom			TLC		
			BLEU	ROUGE-L	METEOR	BLEU	ROUGE-L	METEOR
Average	Baseline	Hybrid-DRL	21.33	27.68	11.09	16.49	30.00	11.88
		CodeTrans	25.25	35.05	15.78	18.30	33.71	14.42
		Re ² Com	23.04	30.58	13.09	19.17	32.37	13.81
		Rencos	24.39	30.77	14.22	20.17	33.01	14.33
		EditSum	25.80	31.02	14.18	18.59	30.80	12.35
		AST-Trans	25.33	36.41	17.30	19.42	29.74	14.61
	Our Approach	DOME	31.83	42.45	20.48	22.20	36.67	16.47
		DOME w/o ISA	30.52	40.09	19.24	21.47	35.61	15.18
		DOME w/o ER	30.09	39.57	18.92	20.58	34.59	14.44

結果

- ISAとERの両コンポーネントはDOMEの性能に寄与している

まとめ

意図を考慮したコメント生成モデルDOMEを提案

- DRPモデルを利用し,コメントの模範解答を取得
- コードと模範解答をエンコーダに入力
- ISAコンポーネントによりノイズを無視

コメント意図ラベル付きデータセットのため自動コメント意図分類器COINを開発

2つのjavaデータセットでDOMEを評価,ベースラインを上回ることを確認

所感

Transformerの仕組みを把握しておらずISAコンポーネントの詳細な動作が理解できなかった

補足 RQ1,2結果

Intent	Method		Funcom			TLC		
			BLEU	ROUGE-L	METEOR	BLEU	ROUGE-L	METEOR
What	Baseline	Hybrid-DRL	22.44	25.89	10.67	19.69	32.89	13.18
		CodeTrans	26.30	32.87	15.05	21.35	36.39	15.63
		Re ² Com	24.17	28.72	12.53	22.21	35.11	14.94
		Rencos	26.19	31.10	14.61	23.28	36.89	16.02
		EditSum	27.58	31.06	14.24	21.34	33.73	13.42
		AST-Trans	27.84	38.28	18.49	23.42	34.24	17.17
	Our Approach	DOME	33.29	41.67	20.53	25.39	39.56	18.22
		DOME <i>w/o ISA</i>	32.01	38.74	18.65	24.37	37.93	16.51
		DOME <i>w/o ER</i>	31.33	38.12	18.57	23.82	37.24	16.37
Why	Baseline	Hybrid-DRL	22.65	27.61	11.14	16.47	28.61	11.56
		CodeTrans	26.52	35.04	15.88	17.58	32.18	13.96
		Re ² Com	24.39	30.75	13.22	18.99	31.45	13.09
		Rencos	24.81	30.12	14.21	20.55	32.99	14.54
		EditSum	27.17	30.98	14.66	18.42	29.85	11.81
		AST-Trans	25.96	35.74	17.77	19.31	29.37	14.94
	Our Approach	DOME	33.07	42.31	20.56	21.97	35.31	15.77
		DOME <i>w/o ISA</i>	31.79	39.37	19.11	21.41	34.27	15.56
		DOME <i>w/o ER</i>	31.13	38.78	18.99	19.60	32.16	12.71
How-to-use	Baseline	Hybrid-DRL	21.16	24.75	10.04	10.25	19.78	7.45
		CodeTrans	25.79	32.91	15.24	13.50	24.26	10.28
		Re ² Com	23.17	27.79	12.18	14.18	23.45	10.09
		Rencos	25.54	27.74	13.34	14.62	22.61	10.21
		EditSum	25.25	27.25	12.79	14.00	21.51	9.07
		AST-Trans	24.93	30.90	15.09	13.24	18.20	9.16
	Our Approach	DOME	31.63	39.31	19.34	17.16	26.11	12.36
		DOME <i>w/o ISA</i>	30.57	37.24	17.54	16.74	25.59	11.25
		DOME <i>w/o ER</i>	30.42	37.15	17.37	15.33	24.96	10.94
How-it-is-done	Baseline	Hybrid-DRL	17.09	25.29	8.50	14.52	29.36	10.02
		CodeTrans	20.65	32.46	13.21	16.36	33.07	12.89
		Re ² Com	18.61	27.89	10.41	17.08	31.04	11.99
		Rencos	19.84	29.28	12.53	18.58	33.73	13.12
		EditSum	22.22	29.73	12.62	16.84	31.18	11.12
		AST-Trans	19.65	33.60	14.40	17.61	30.32	13.29
	Our Approach	DOME	26.98	39.52	17.65	20.48	36.66	14.73
		DOME <i>w/o ISA</i>	26.03	38.19	18.20	19.50	36.57	13.10
		DOME <i>w/o ER</i>	25.78	37.73	18.10	19.14	35.72	13.01
Property	Baseline	Hybrid-DRL	23.30	34.84	15.09	21.53	39.37	17.17
		CodeTrans	27.00	41.95	19.53	22.70	42.63	19.33
		Re ² Com	24.85	37.77	17.08	23.37	40.79	18.92
		Rencos	25.57	35.60	16.39	23.82	38.85	17.76
		EditSum	26.79	36.09	16.60	22.35	37.74	16.33
		AST-Trans	28.29	43.54	20.73	23.54	36.56	18.47
	Our Approach	DOME	34.18	49.43	24.32	26.01	45.73	21.29
		DOME <i>w/o ISA</i>	32.21	46.92	22.70	25.35	43.68	19.47
		DOME <i>w/o ER</i>	31.80	46.09	21.59	25.01	42.85	19.16
Average	Baseline	Hybrid-DRL	21.33	27.68	11.09	16.49	30.00	11.88
		CodeTrans	25.25	35.05	15.78	18.30	33.71	14.42
		Re ² Com	23.04	30.58	13.09	19.17	32.37	13.81
		Rencos	24.39	30.77	14.22	20.17	33.01	14.33
		EditSum	25.80	31.02	14.18	18.59	30.80	12.35
		AST-Trans	25.33	36.41	17.30	19.42	29.74	14.61
	Our Approach	DOME	31.83	42.45	20.48	22.20	36.67	16.47
		DOME <i>w/o ISA</i>	30.52	40.09	19.24	21.47	35.61	15.18
		DOME <i>w/o ER</i>	30.09	39.57	18.92	20.58	34.59	14.44