Crash Report Analysis and Classification

Presenter: Wen Cui

About Crash Report Prioritization

Crash Report Overview

- Large amount of crash reports
- First come, first served may delay fix of important crashes
- Sometimes, prior knowledge is not enough

Need of Crash Report Prioritization Tools

To build the gap, this project

S

STUDY

Characteristics of crash report

A

CONSTRUCT

Use Few-shot learning, Similarity match, CNN for crash report classification

C

COMPARISON

Compare crash report classification tools

P

PROPOSE

Propose future direction of crash report classification

About Crash Report

Crash ID: 7aa331bd-5c83-4951-8c4f-f86720230323

Signature: [@ AsyncShutdownTimeout | profile-change-teardown | ServiceWorkerShutdownBlocker: shutting down Service Workers]

Details | Crash Annotations | Bugzilla | Modules | Raw Data and Minidumps | Extensions | Telemetry Environment | Correlations | Debug

Signature	AsyncShutdownTimeout | profile-change-teardown | ServiceWorkerShutdownBlocker: shutting down Service Workers **More Reports** **Search**
UUID	7aa331bd-5c83-4951-8c4f-f86720230323
Date Processed	2023-03-23 19:37:52 UTC
Uptime	265,740 seconds (3 days, 1 hour and 49 minutes)
Last Crash	17,444,199 seconds before submission (28 weeks, 5 days and 21 hours)
Install Age	265,740 seconds since version was first installed (3 days, 1 hour and 49 minutes)
Install Time	2023-03-20 17:28:54
Product	Firefox
Release Channel	nightly
Version	113.0a1
Build ID	20230315092641 (2023-03-15) **Buildhub data**
OS	macOS 13
OS Version	13.0.1 22A400
Build Architecture	arm64
CPU Info	
CPU Count	10
Adapter Vendor ID	0x106b
Adapter Device ID	
Startup Crash	False
Process Type	parent
MOZ_CRASH Reason (Sanitized)	[Parent 19736, Main Thread] ###!!! ABORT: file /builds/worker/checkouts/gecko/dom/serviceworkers/ServiceWorkerShutdownBlocker.cpp:110
Crash Reason	EXC_BAD_ACCESS / KERN_INVALID_ADDRESS
Crash Address	0x0000000000000000
Available Physical Memory	18,635,128,832 bytes (18.64 GB)
EMCheckCompatibility	True
App Notes	FP(D00-L1000-W0000000-T010) WR? WR+ GL Context? GL Context+ WebGL? WebGL+ xpcom_runtime_abort(###!!! ABORT: file /builds/worker/checkouts/gecko/dom/ser

Feature Extraction

System-related	Total Physical memory, thread count, processor notes, CPU count
Crash-related	Method signature, prior fixes. Startup crash, module count
Other	Crash type, last crash, frame count

Crash Report Process

Apply classification Algorithms

CRASH

Firefox crashes

REPORT

Failing stack trace collected

CLASSIFY

Classify automatically generated crash report

FIX

Bug fixed

CLASSIFICATION APPROACHES

Existing Techniques

Idea: Convert crash report to sentences and perform text classification.

Not suitable for small amount of training data
Not-convincing definition "top crashes"

Machine Learning

New Approaches

Few shot Learning
Similarity Match
CNN

RESEARCH QUESTIONS

RQ1	RQ2
How do we classify crash reports when there is little training data?	How does few-shot learning perform compared with other approaches in terms of crash report classification?

RQ1: DATA RETRIEVAL

STEP1: Data Collection

Mozilla Crash Report

Collected data for consecutive 30 days

Filter and Assign Labels

```
OOM | large | js::AutoEnterOOMUnsafeRegion::crash |
js::AutoEnterOOMUnsafeRegion::crash | JS::CallbackTracer::onEdge 8249831424
30>>> Start processing: 2023-03-06 00:37:34.170139+00:00
(processor_ip-172-31-2-221_us-west-2_compute_internal_8);8;597;2802;39;32,1
```

Assign label based on occurrence and fix time

Feature Extraction

Compact crash report to text

RQ1: DATA RETRIEVAL

STEP2: Create training and test set

Select crash report with labels

cnn_train_2.csv

```
text,label
js::gc::HeaderWord::get 8587350016 30>>> Start processing: 2023-03-06
00:26:08.547884+00:00 (processor_ip-172-31-24-89_us-
west-2_compute_internal_8);6;281;106;39;45,1
js::ObjectGroup::sweep 6442450944 34>>> Start processing: 2023-03-02
00:08:39.963803+00:00 (processor_ip-172-31-23-172_us-
west-2_compute_internal_8);2;124;774942;275;27,1
<unknown in SHCore.dll> | CDeviceBase::_DevQueryCallback 4149985280
67>>> Start processing: 2023-03-10 03:33:08.718026+00:00
(processor_ip-172-31-32-95_us-
west-2_compute_internal_8);2;62;2169;157;14140344;10,1
OOM | small 1721098240 90>>> Start processing: 2023-03-02
00:22:42.895393+00:00 (processor_ip-172-31-33-118_us-
west-2_compute_internal_8);2;11060;68720;160;396262;35,1
mozilla::dom::quota::QuotaManager::Shutdown::<T>::operator()
17128787968 63>>> Start processing: 2023-02-26 00:55:15.548001+00:00
(processor_ip-172-31-17-241_us-
west-2_compute_internal_7);8;2795;231;113;43068027;22,0
```

Collect new crash data without labels

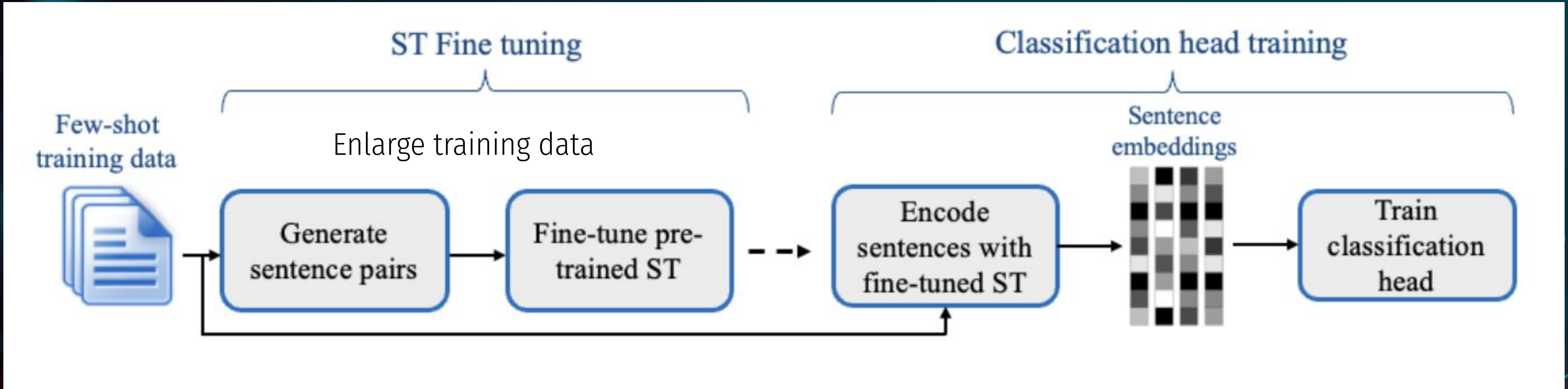
cnn_test.csv — Edited

```
text
shutdownhang |
js::frontend::ExtensibleCompilationStencil::~ExtensibleCompilationSte
ncil 4171517952 52>>> Start processing: 2023-03-07
10:40:04.558791+00:00 (processor_ip-172-31-7-120_us-
west-2_compute_internal_9);4;136;1061;116;31277933;7
shutdownhang | NtQueryVirtualMemory 5259399168 35>>> Start
processing: 2023-02-26  00:12:36.934910+00:00
(processor_ip-172-31-29-60_us-
west-2_compute_internal_8);4;87;107;97;87698;7
mozilla::dom::quota::QuotaManager::Shutdown::<T>::operator()
8483495936 73>>> Start processing: 2023-03-10 03:43:58.728717+00:00
(processor_ip-172-31-11-157_us-
west-2_compute_internal_8);4;2795;14986;156;35002;22
AsyncShutdownTimeout | IOUtils: waiting for profileBeforeChange IO to
complete | JSON store: writing data for 'targeting.snapshot'
8527294464 60">>> Start processing: 2023-03-06 00:38:02.761447+00:00
(processor_ip-172-31-11-157_us-west-2_compute_internal_8)
SignatureShutdownTimeout: Signature replaced with a Shutdown Timeout
signature; was: ""Abort | NS_DebugBreak | nsDebugImpl::Abort |
XPTC__InvokebyIndex""";4;1047;423247;178;16
sys_read 3174739968 28>>> Start processing: 2023-03-07
10:34:16.647637+00:00 (processor_ip-172-31-32-95_us-
west-2_compute_internal_8) mdsw did not identify the crashing
thread";2;332;80756;137;0
```

RQ1: APPROACH DESCRIPTION

SETFIT

SETFIT (Sentence Transformer Fine-tuning), an efficient and prompt-free framework for few-shot fine-tuning of Sentence Transformers (ST)



RQ1: APPROACH DESCRIPTION

Few-shot Learning

1. Load Training dataset

```
dataset = load_dataset('csv', data_files={
    'train': ['data/train_2.csv'],
    'eval': ['data/eval_2.csv']},
    cache_dir="./data/"
)
```

2. Load a SetFit model from Hub

```
model = SetFitModel.from_pretrained(
    "sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2",
    cache_dir="./models/"
)
```

RQ1: APPROACH DESCRIPTION

Few-shot Learning

3. Create Trainer

```
trainer = SetFitTrainer(
    model=model,
    train_dataset=dataset['train'],
    eval_dataset=dataset['eval'],
    loss_class=CosineSimilarityLoss,
    metric="accuracy",
    batch_size=16,
    num_iterations=20,  # The number of text pairs to generate for contrastive learning
    num_epochs=1,  # The number of epochs to use for contrastive learning
    column_mapping={"text": "text", "label": "label"}  # Map dataset columns to text/label expected by trainer
)
```

RQ1: APPROACH DESCRIPTION

Few-shot Learning

4. Train, Evaluate, Save

```
# Train and evaluate
trainer.train()
metrics = trainer.evaluate()

# save
trainer.model._save_pretrained(save_directory="./output/")
```

5. Inference

Preds = model(test_list)

RQ1: APPROACH DESCRIPTION

SIMILARITY MATCH

Levenshtein distance → Measure string difference: min single-character edits required to change one word into the other

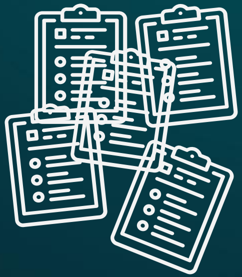
$$\operatorname{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \operatorname{lev}_{a,b}(i-1,j) + 1 \\ \operatorname{lev}_{a,b}(i,j-1) + 1 \\ \operatorname{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

RQ1: APPROACH DESCRIPTION

SIMILARITY MATCH

Income Crash Report

Set of Crash Report with Known Classification



Compute levenshtein distance

Enlarge training data

Similarity Score

Assign ranking for highest score

RQ1: APPROACH DESCRIPTION

CNN MODEL

Create
Train-validation split

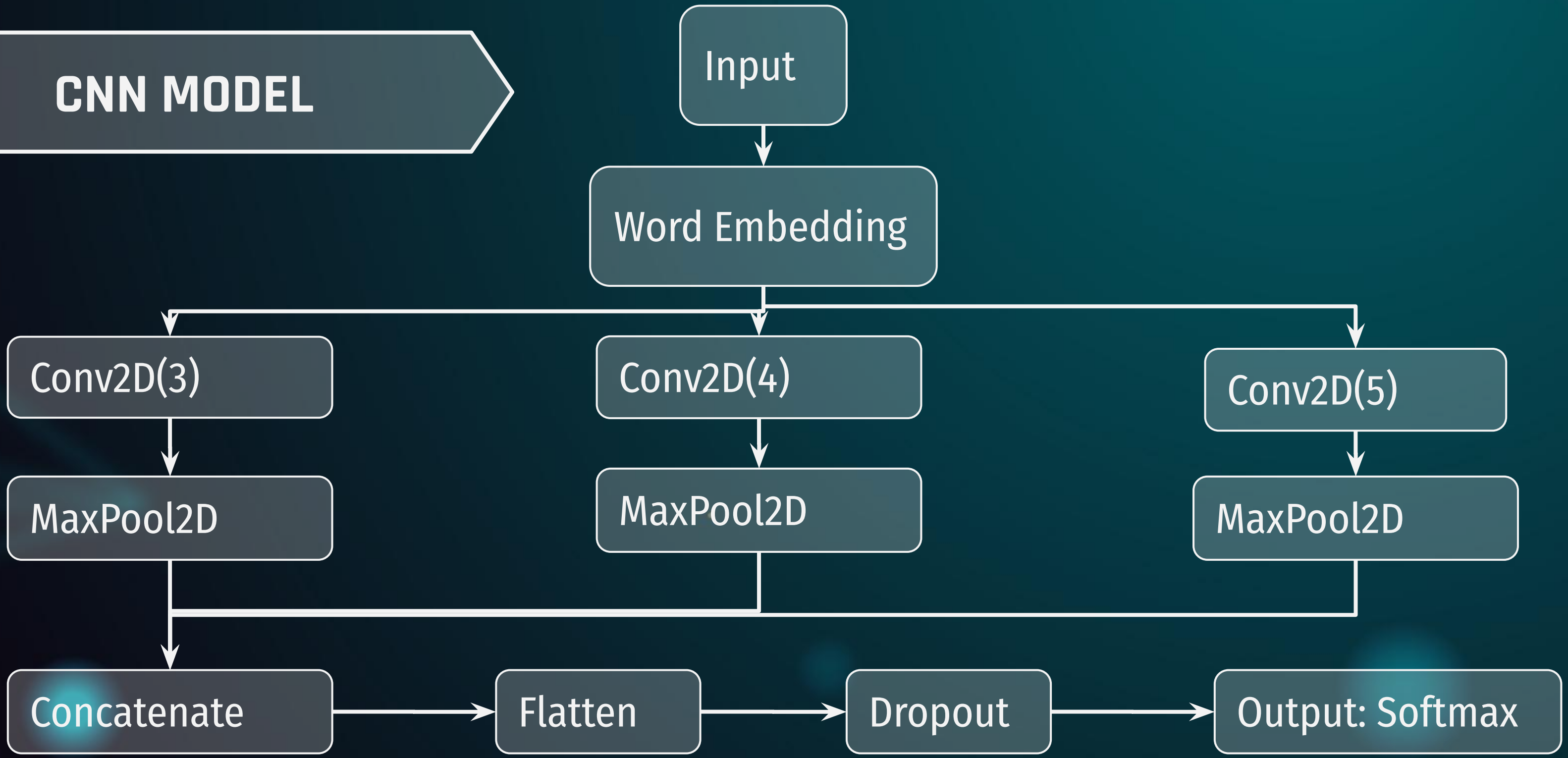
Create embedding matrix

Create the model

Train the model

RQ1: APPROACH DESCRIPTION

CNN MODEL



RQ2: APPROACH COMPARISON

Step

01

Create two sets of train-test set

02

Run both sets using 3 approaches

03

Output Comparison

DATA COLLECTION

Two sets of training-testing data

Each set contains 5 training reports with 10 ✖ 5 test reports

RQ2: RESULTS

set_1	expected_output	fs_output	fs_accuracy	sim_output	sim_accuracy	cnn_output	cnn_accuracy
test_1	0 0 0 1 0	1 0 0 0 0	0.6	1 1 0 0 0	0.4	1 1 1 1 1	0.2
test_2	0 1 0 1 0	1 1 0 1 0	0.8	1 0 0 0 0	0.4	0 1 1 1 1	0.6
test_3	1 1 1 0 0	0 1 0 0 0	0.6	0 0 1 0 1	0.4	0 0 1 1 1	0.2
test_4	1 0 1 1 1	1 0 1 1 0	0.8	0 1 1 0 0	0.2	1 0 0 1 1	0.6
test_5	1 1 1 1 0	1 0 1 1 0	0.8	1 0 0 1 0	0.6	1 0 1 1 1	0.6
test_6	0 1 0 0 1	1 1 0 1 1	0.6	0 0 0 0 0	0.6	1 0 1 1 1	0.2
test_7	0 1 0 1 1	0 1 0 0 1	0.8	1 0 0 0 1	0.4	1 1 0 1 1	0.8
test_8	1 1 0 1 0	1 1 0 1 0	1	1 0 0 0 0	0.6	0 1 1 1 1	0.4
test_9	1 1 1 1 1	1 1 0 1 0	0.6	0 0 0 1 0	0.2	1 1 1 1 0	0.8
test_10	1 1 1 1 0	1 0 1 0 1	0.4	0 0 0 0 0	0.2	1 1 1 1 0	1
AVERAGE			0.7		0.4		0.54

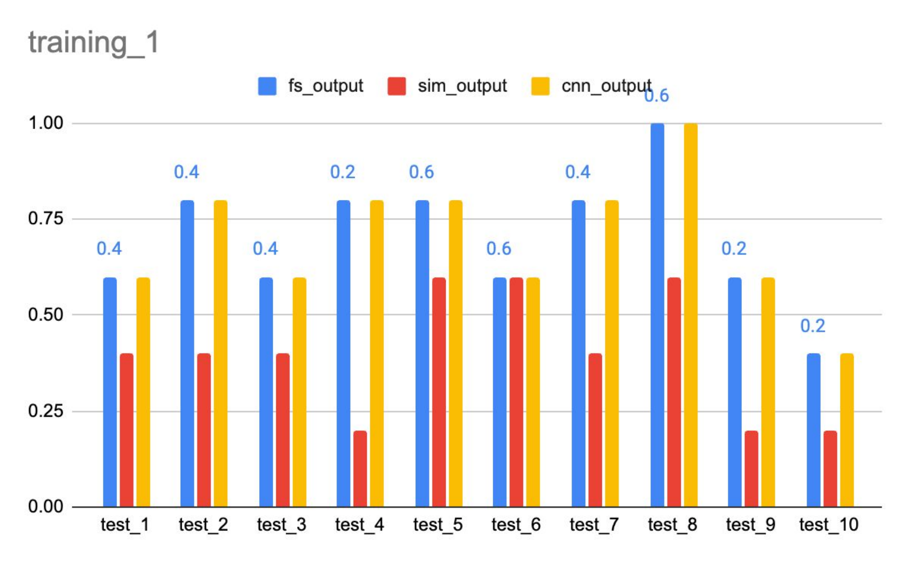
Training set #1 → Few shot > CNN > Similarity Match

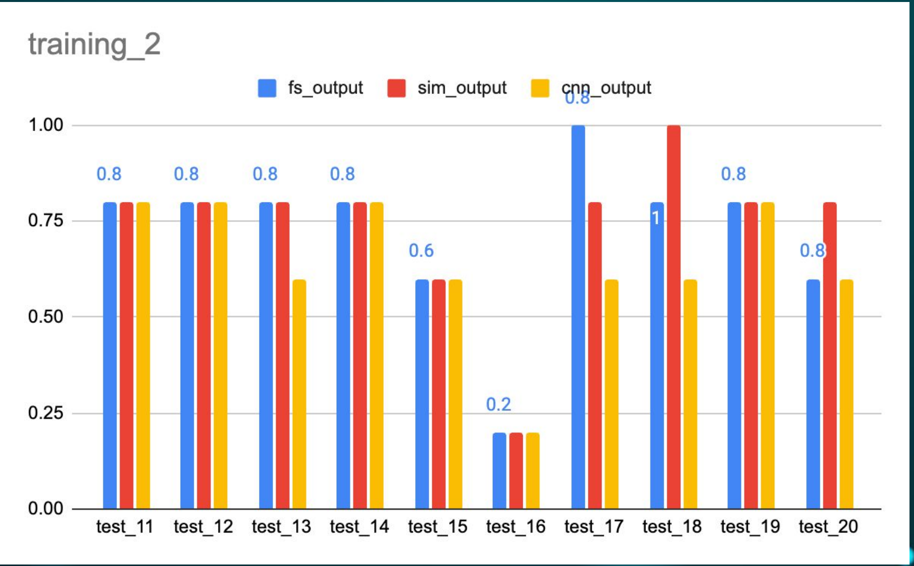
RQ2: RESULTS

set_2	expected_output	fs_output	fs_accuracy	sim_output	sim_accuracy	cnn_output	cnn_accuracy
test_11	0 1 1 1 1	1 1 1 1 1	0.8	1 1 1 1 1	0.8	1 1 1 1 1	0.8
test_12	1 1 1 0 1	1 1 1 1 1	0.8	1 1 1 1 1	0.8	1 1 1 1 1	0.8
test_13	0 0 1 1 0	1 0 1 1 0	0.8	0 1 1 1 0	0.8	1 1 1 1 1	0.6
test_14	0 0 0 1 1	0 0 1 1 1	0.8	0 0 1 1 1	0.8	0 0 1 1 1	0.8
test_15	1 0 1 0 1	1 1 1 1 1	0.6	1 1 1 1 1	0.6	1 1 1 1 1	0.6
test_16	0 1 0 0 0	1 1 1 1 1	0.2	1 1 1 1 1	0.2	1 1 1 1 1	0.2
test_17	1 0 1 1 1	1 0 1 1 1	1	1 0 1 0 1	0.8	1 1 1 0 1	0.6
test_18	0 1 0 1 1	1 1 0 1 1	0.8	0 1 0 1 1	1	1 1 1 1 1	0.6
test_19	0 1 0 1 1	1 1 0 1 1	0.8	1 1 0 1 1	0.8	1 1 0 1 1	0.8
test_20	0 1 1 0 1	1 1 1 1 1	0.6	0 1 1 1 1	0.8	1 1 1 1 1	0.6
			0.72		0.74		0.64

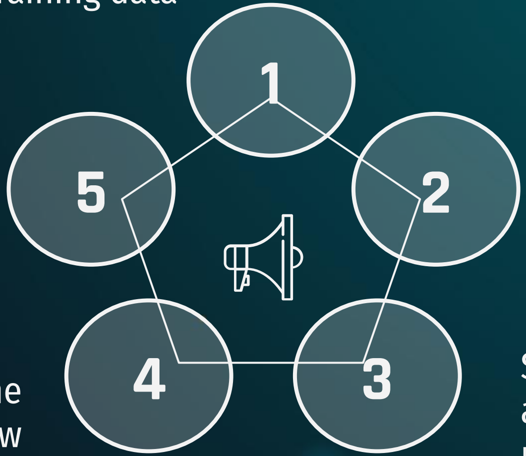
Training set #1 → Few shot ~ Similarity Match > CNN

RQ2: RESULTS


training_1
fs_output
sim_output
cnn_output
1.00
0.75
0.50
0.25
0.00
0.4
0.4
0.4
0.2
0.6
0.6
0.4
0.6
0.2
0.2
test_1
test_2
test_3
test_4
test_5
test_6
test_7
test_8
test_9
test_10


training_2
fs_output
sim_output
cnn_output
1.00
0.75
0.50
0.25
0.00
0.8
0.8
0.8
0.8
0.6
0.2
0.8
1
0.8
0.8
test_11
test_12
test_13
test_14
test_15
test_16
test_17
test_18
test_19
test_20

FINDINGS



1. Overall, few shot learning outperforms the other two approaches with less training data
2. Few shot learning performance could be improved by hyper-parameter tuning
3. Similarity-based approach performance highly depends on training data quality
4. CNN does not work well when the amount of training data is low
5. More data could be used to make results more convincing

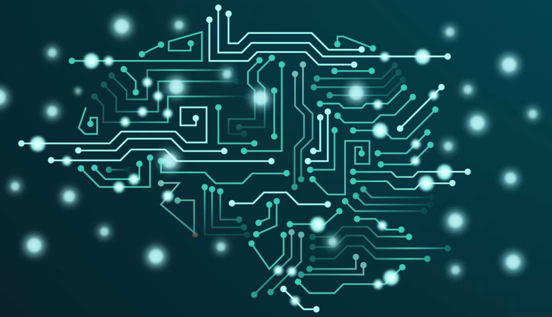
CHALLENGES

DATA COLLECTION

Difficulty in collecting old data

REPORT ANALYSIS

Difficulty in finding resources of crash report features



APPROACH COMPARISON

Difficulty in finding other classification techniques

GENERALIZABI LITY

Hard to illustrate generalizability of data due to short amount of training data

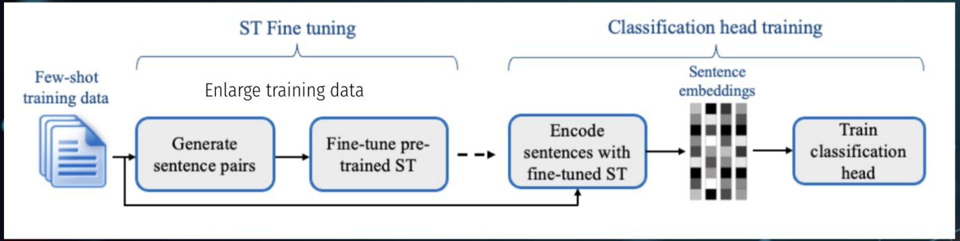
WHAT TO DO NEXT ...

Crash Report classification techniques → Improve few shot learning performance

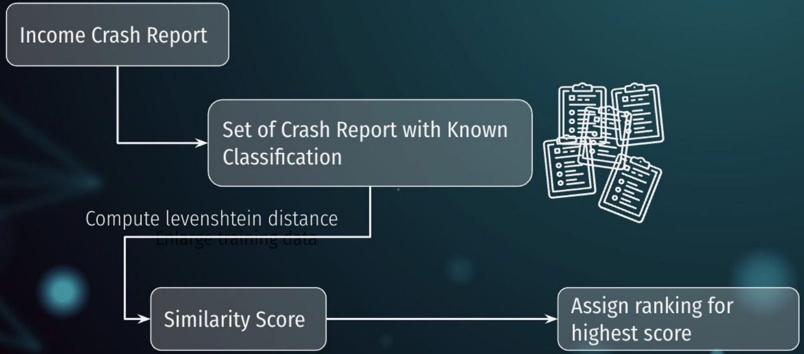
Technique comparison → Investigate into more approaches and use more data for comparison

SETFIT

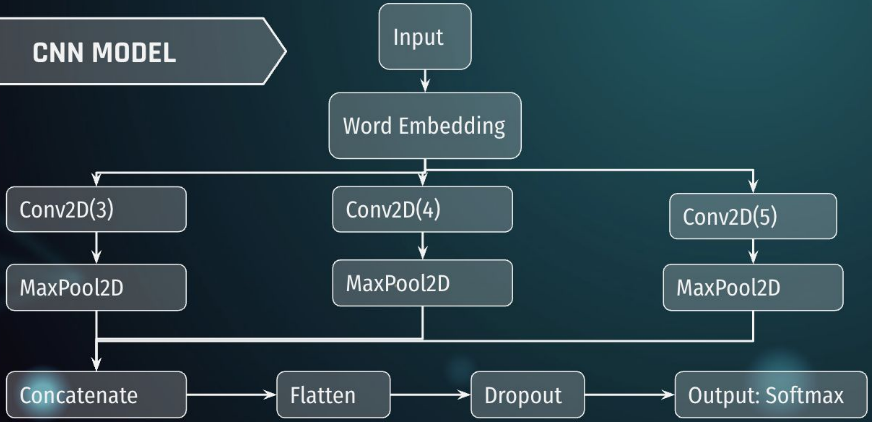
SETFIT (Sentence Transformer Fine-tuning), an efficient and prompt-free framework for few-shot fine-tuning of Sentence Transformers (ST)



SIMILARITY MATCH



CNN MODEL



Overall, few shot learning outperforms the other two approaches with less training data

Few shot learning performance could be improved by hyper-parameter tuning

Similarity-based approach performance highly depends on training data quality

CNN does not work well when the amount of training data is low

More data could be used to make results more convincing