

TFix: Learning to Fix Coding Errors with a Text-to-Text Transformer

Berkay Berabi, Jingxuan He, Veselin Raychev, Martin Vechev

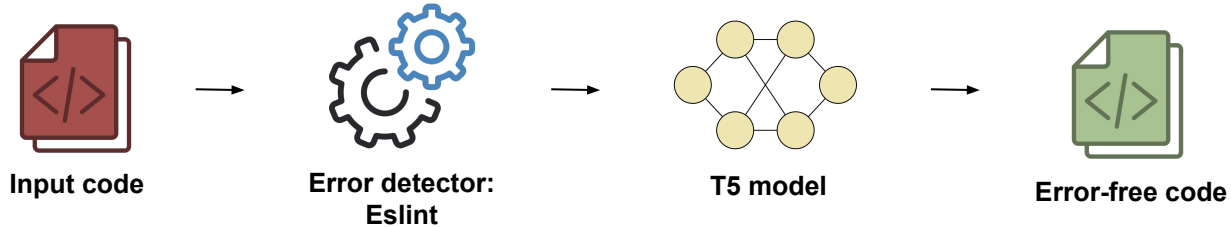
Contact: berkay.berabi@gmail.com

ETH zürich

 **SRILAB**



Overview: General Pipeline



```
if (value != null && fieldData.type != null) {  
  var type = null  
  for (var typeEntry in types) {  
    var typeNames = types(typeEntry);  
    if(typeNames.indexOf(fieldData.type) >= 0)  
  {  
    type = typeEntry;
```

```
var type = null  
for (var typeEntry in types) {  
  if(!types.hasOwnProperty(typeEntry)) continue;  
  var typeNames = types(typeEntry);
```

Input: fix <error type> <error message> <error line>: <error context>

Fine-tuning

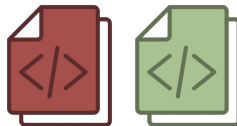
Dataset

~5.5 millions of GitHub commits



Dataset
construction

Extracted code patch pairs
(erroneous, fixed)



Final Dataset: ~100k samples with 52 error types

Advantages

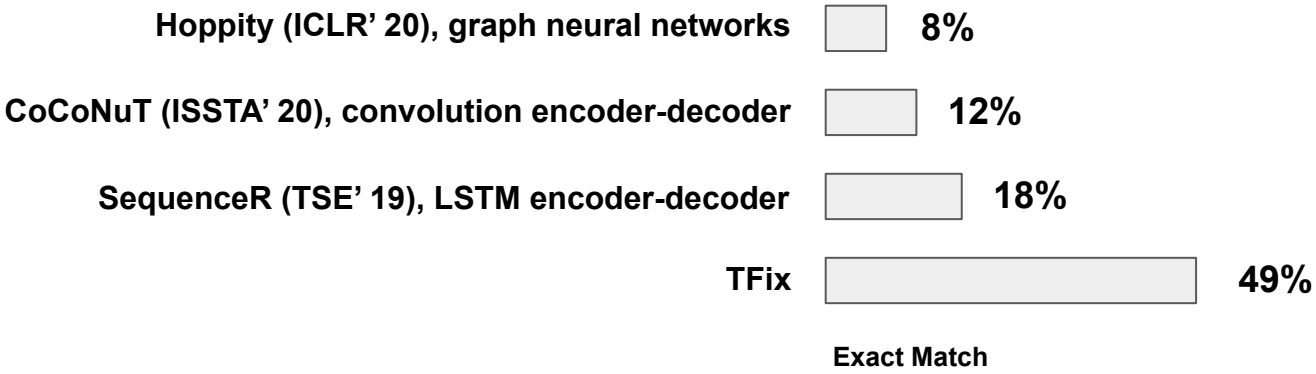
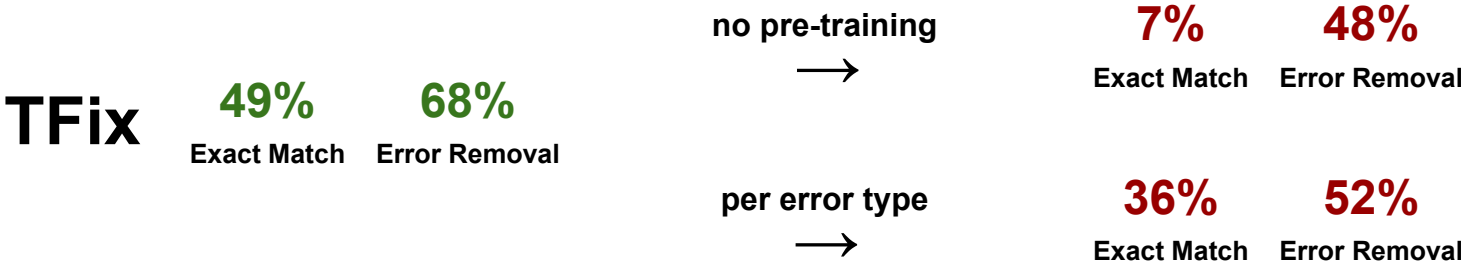
Usage of T5 model pre-trained on natural languages

→ Transfer learning between code and natural languages

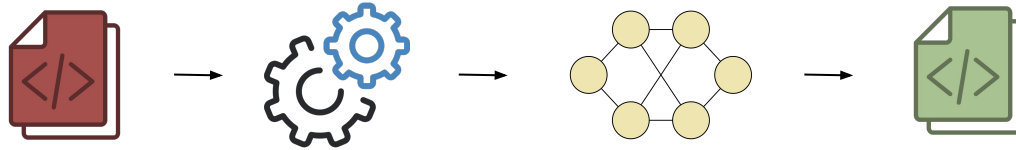
Fine-tuning all error types together

→ Data augmentation and multi-task learning across various error types

Evaluation



Summary



Code, data, and models: <https://github.com/eth-sri/TFix>