How Do Python Framework APIs Evolve?

An Exploratory Study

Zhaoxu Zhang, Hengcheng Zhu, Ming Wen, Yida Tao,

Yepang Liu, and Yingfei Xiong

Related Work

- Xavier et al. **Historical and Impact Analysis of API Breaking Changes: A Large-scale Study**. [SANER 2017]

- Wei et al. **Taming Android Fragmentation: Characterizing and Detecting Compatibility Issues for Android Apps** . [ASE 2016]

- Meng et al. **A History-based Matching Approach to Identification of Framework Evolution**. [ICSE 2012]

- Dig et al. **How do APIs Evolve? A Story of Refactoring**. [*Journal of Software Maintenance 2006*]

...

Why Python

- Lack of type information in code

- Type checking is deferred to run time

- Argument passing is flexible
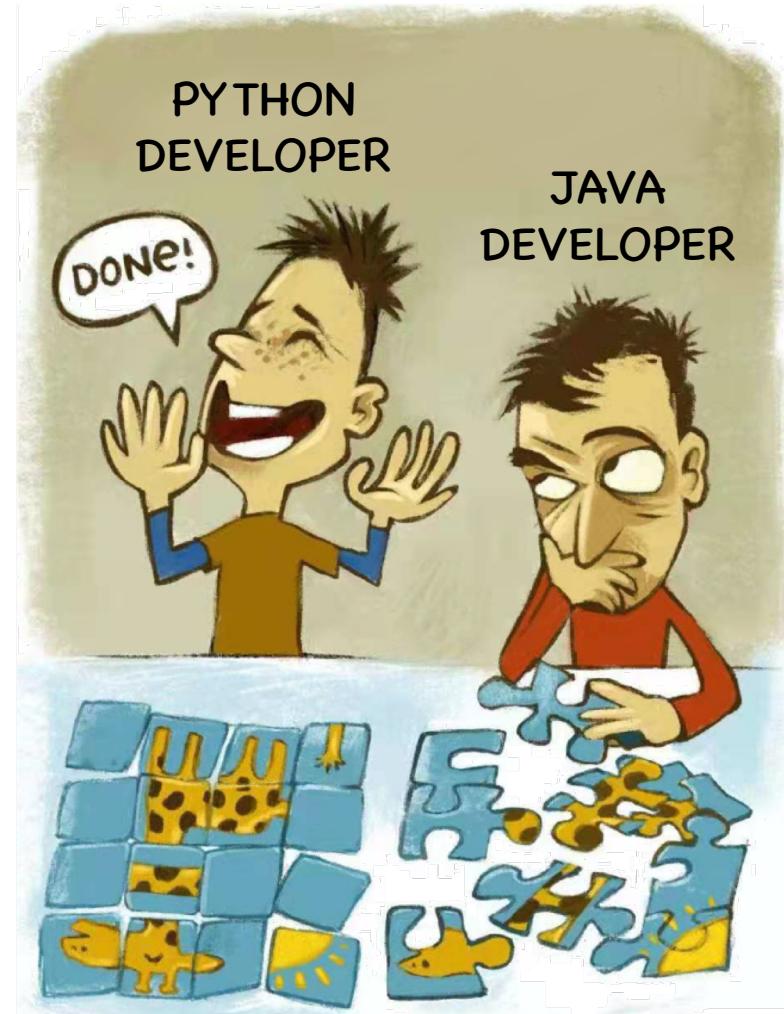
Special Issues in Python

OneHotEncoder ,TypeError: __init__() got an unexpected keyword argument 'drop'

Asked 8 months ago   Active 6 months ago   Viewed 2k times

In sklearn.learn...

1   has p

I didn

seen

Colu

Futur

tensorflow:AttributeError: 'module' object has no attribute 'mul'

Asked 2 years, 11 months ago   Active 3 days ago   Viewed 60k times

```
Traceback (most r
    File "<stdin>",
AttributeError: '
>>> enter code he
```

Accord

Django error: render_to_response() got an unexpected keyword argument 'context_instance'

Asked 3 years, 6 months ago   Active 1 year, 4 months ago   Viewed 30k times

178

```
tf.m
tf.s
```

Exception T
Exception Va

The `context`

You'll r

Django 1.8,

The solution

RequestCont

pandas DataFrame.rename unexpected keyword argument "axis" when using mapper

Asked 2 years, 1 month ago   Active 2 years, 1 month ago   Viewed 10k times

Following the pandas docs I tried the following (verbatim out of the docs):

7

```
df = pd.DataFrame({"A": [1, 2, 3], "B": [4, 5, 6]})
df.rename(str.lower, axis='columns')
```

yet I'm getting the error

What We Explored

How do **Python** framework APIs **evolve**?

What are the **compatibility issues**?

How do developers **fix** these issues?

Empirical Study

Python framework APIs evolution

Subject Selection

Framework	Category	# Releases	Versions	# Projects
TensorFlow 2.0	Deep learning	68	`0.12.0rc0 − 2.0.0rc2`	1,000
K Keras		35	`1.0.2 − 2.2.4`	1,000
scikit learn	Data analytics	24	`0.15.0b1 − 0.20.3`	883
pandas $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$		42	`0.9.1 − 0.24.2`	1,000
Flask web development, one drop at a time	Web development	12	`0.10 − 1.0.2`	1,000
django		108	`1.8 − 2.2rc1`	1,000
Total		**288 releases**		**5,538**

Subject Selection

6 Frameworks
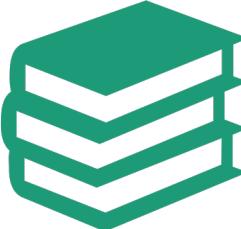
3 Categories

288 Releases

5,538 Projects

Dataset Collection

Evolution Patterns

6 Frameworks (TensorFlow 2.0, Keras, pandas, django, scikit-learn, Flask)

Dynamic Extraction → Framework APIs → Classification → 1,094,359 API Changes In 14 types

Issue & Fixing

5,538 GitHub Projects → Analyze Issue Reports and Commits → 409 Issues + 347 Fixes
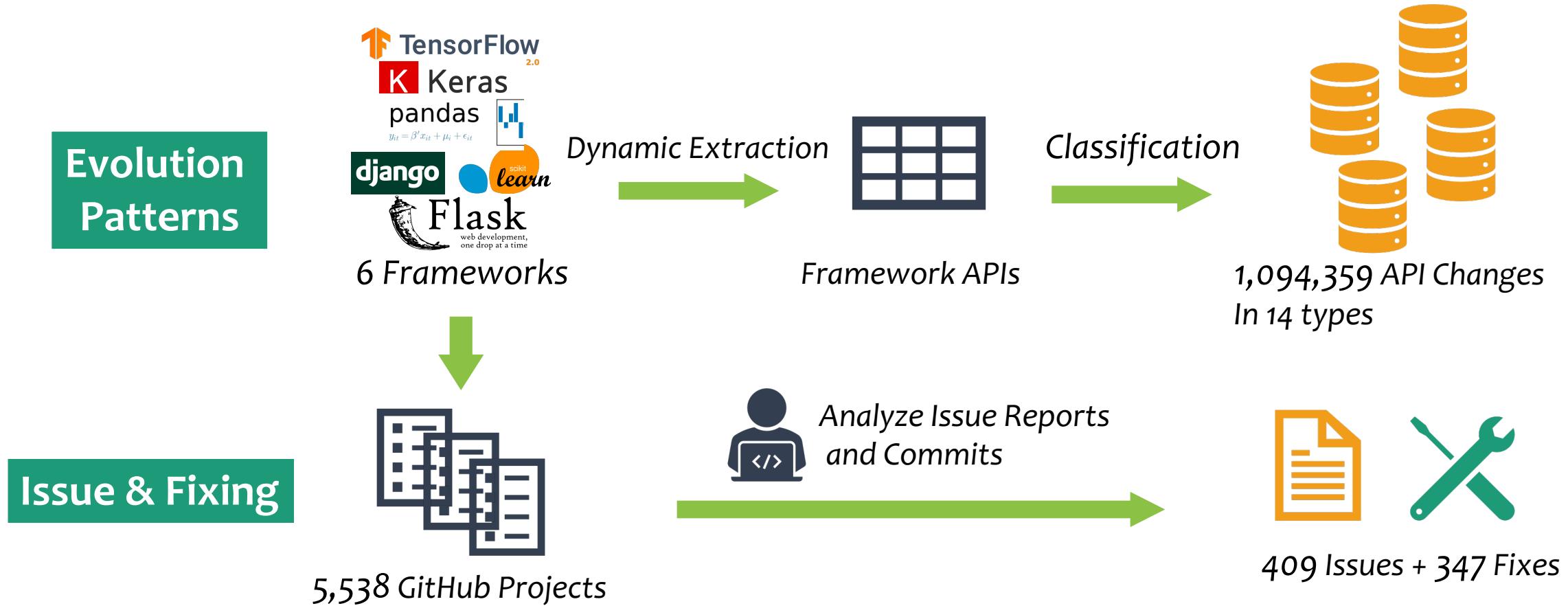
Evolution Patterns

TABLE II
API EVOLUTION PATTERNS AND MATCHING RULES

Index	Element	Pattern	Frequency	Matching Rule	Change Type
1	Class (c)	Class Removal*	25,481 (2.3%)	$c \in C \wedge c \notin C'$	Breaking
2		Class Addition*	26,694 (2.4%)	$c \notin C \wedge c \in C'$	Non-Breaking
3	Method (m)	Method Removal*	309,759 (28.3%)	$m \in M \wedge m \notin M'$	Breaking
4		Method Addition*	319,618 (29.2%)	$m \notin M \wedge m \in M'$	Non-Breaking
5		Required Parameter Addition*	5,026 (0.5%)	$\exists p \notin P, p \in P' \wedge \neg\text{opt}(p)$	Breaking
6		Required Parameter Removal*	3,218 (0.3%)	$\exists p \in P, p \notin P' \wedge \neg\text{opt}(p)$	Breaking
7		Parameter Reordering*	840 (0.1%)	$\exists p_1, p_2 \in P \cap P', p_1 \prec p_2 \wedge p_1' \succ p_2'$	Breaking
8		Optional Parameter Addition	12,996 (1.2%)	$\exists p \notin P, p \in P' \wedge \text{opt}(p)$	Breaking
9		Optional Parameter Removal	4,063 (0.4%)	$\exists p \in P, p \notin P' \wedge \text{opt}(p)$	Breaking
10		Parameter Default Value Addition	1,006 (0.1%)	$\exists p \in P \cap P', \neg\text{opt}(p) \wedge \text{opt}(p')$	Breaking
11		Parameter Default Value Removal	851 (0.1%)	$\exists p \in P \cap P', \text{opt}(p) \wedge \neg\text{opt}(p')$	Breaking
12		Parameter Default Value Change	11,362 (1.0%)	$\exists p \in P \cap P', \text{def}(p) \neq \text{def}(p')$	Breaking
13	Field (f)	Field Removal*	182,826 (16.7%)	$f \in F \wedge f \notin F'$	Breaking
14		Field Addition*	190,619 (17.4%)	$f \notin F \wedge f \in F'$	Non-Breaking

"p" and "P" denote a single parameter and a parameter list, respectively. "c" and "C" denote a single class and a set of classes in a release, respectively ("m" and "M", "f" and "F" can be interpreted similarly). "*" means that this pattern also applies to Java. "$'$" denotes the new API version. "$\text{opt}(p)$" means that the parameter p is optional. "$\text{def}(p)$" denotes the default value of p. "$a \prec b$" / "$a \succ b$" denotes parameter a is before/after b.

API Evolution Patterns in Java [SANER'17]

28.0%

Breaking changes in 501,645 API changes

2.5%

Clients impacted by breaking changes

TABLE I
EVALUATED BREAKING AND NON-BREAKING CHANGES.

Category	API Element	List of Changes
Breaking	Type	REMOVAL, VISIBILITY LOSS, SUPERTYPE CHANGE
	Field	REMOVAL, VISIBILITY LOSS, TYPE CHANGE, DEFAULT VALUE CHANGE
	Method	REMOVAL, VISIBILITY LOSS, RETURN TYPE CHANGE, PARAMETER LIST CHANGE, EXCEPTION LIST CHANGE
Non-breaking	All	ADDITION, VISIBILITY GAIN, DEPRECATION OPERATION

Xavier et al. **Historical and Impact Analysis of API Breaking Changes: A Large-scale Study.** SANER 2017

Impact of API Evolutions in Python

50.5%

Breaking changes in 1,064,081API changes

53.6%

Clients impacted by breaking changes

Breaking Changes ■ Non-breaking Changes

	Breaking Changes	Non-breaking Changes
Python Class	47.70%	52.30%
Java Class	18.92%	81.08%
Python Method	48.30%	51.70%
Java Method	27.81%	72.19%
Python Field	40.90%	59.10%
Java Field	37.41%	62.59%

0%   20%   40%   60%   80%   100%
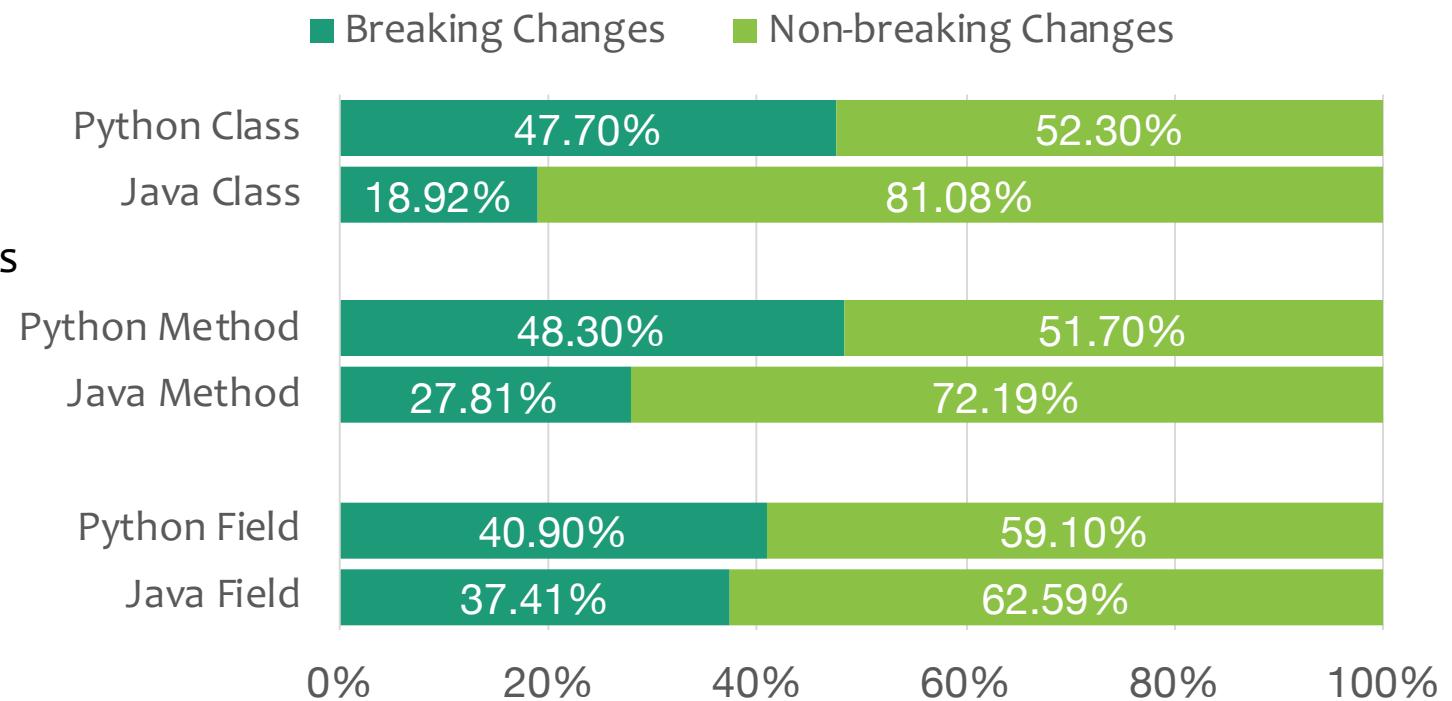
Evolution Patterns (cont.)
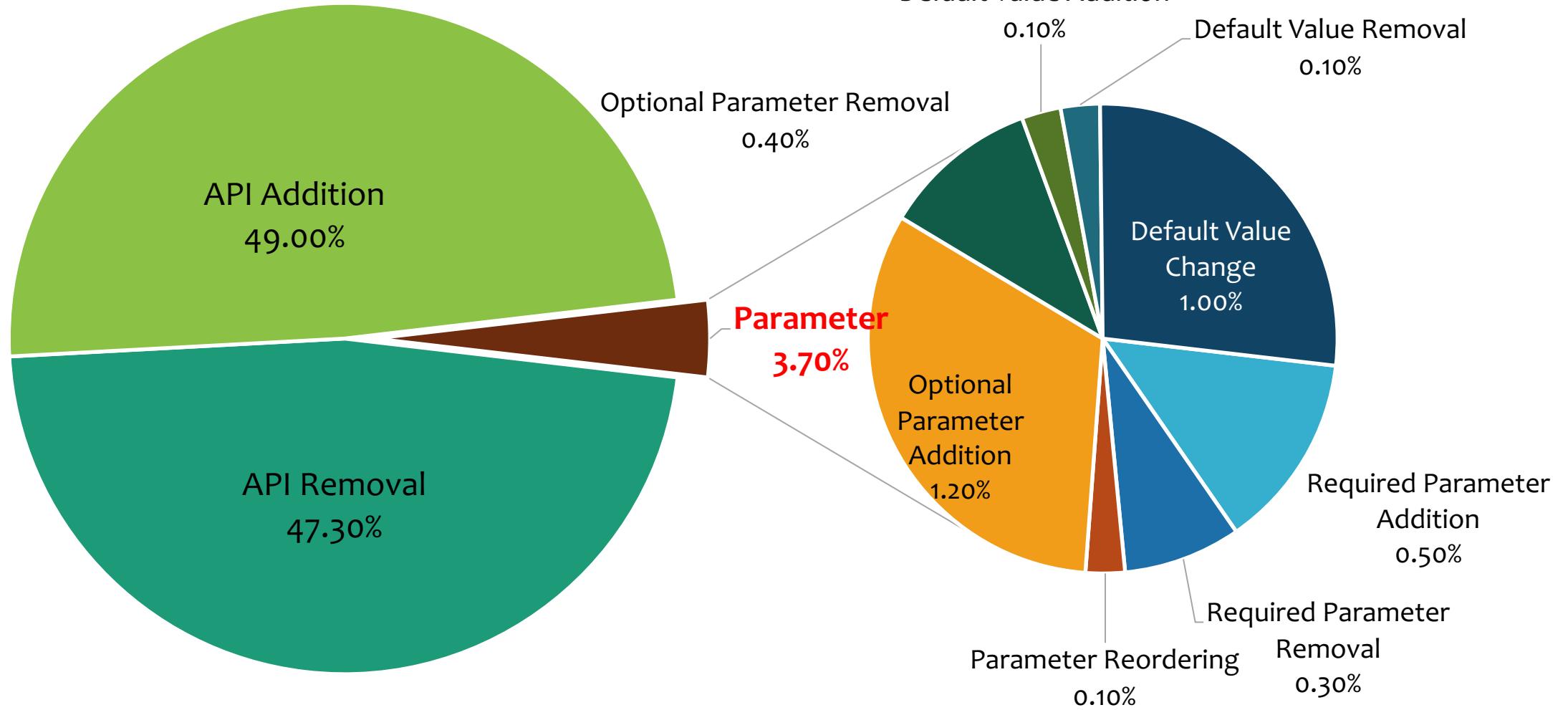
Compatibility Issues (API Existence)

mgeorgiadis commented on May 3, 2019

Hello,

I'm running into issues running pandastable with the latest version of pandas (>0.24). When I try to import pandas table with

from pandastable import Table, TableModel

I get th

```
from pandas.tools import plotting
```

```
from pandastable import Table, TableModel
```

```
ModuleNotFoundError: No module named 'pandas.tools'
```

```
File "C:\Python\WPy64-3720\python-3.7.2.amd64\lib\site-packages\pandastable\core.py", line 42,
    from .plotting import MPLBaseOptions, PlotViewer
File "C:\Python\WPy64-3720\python-3.7.2.amd64\lib\site-packages\pandastable\plotting.py", line
    from pandas.tools import plotting
ModuleNotFoundError: No module named 'pandas.tools'
```

API Evolution Process

```
# Pandas 0.20.0
  pandas.tools. plotting(...)

  pandas.
```

API Addition

API Relocation

```
# Pandas 0.24.0
  pandas.tools.plotting(...)
  pandas.plotting(...)
```

API Removal

Compatibility Issues (cont.)

Issue 133, pandastable

```
from pandas.tool import plotting
```

ImportError (14.2%)

AttributeError (48.9%)

```
pandas.tool.plotting()
```

Compatibility Issues (Parameter Change)

commented on Jun 13, 2017
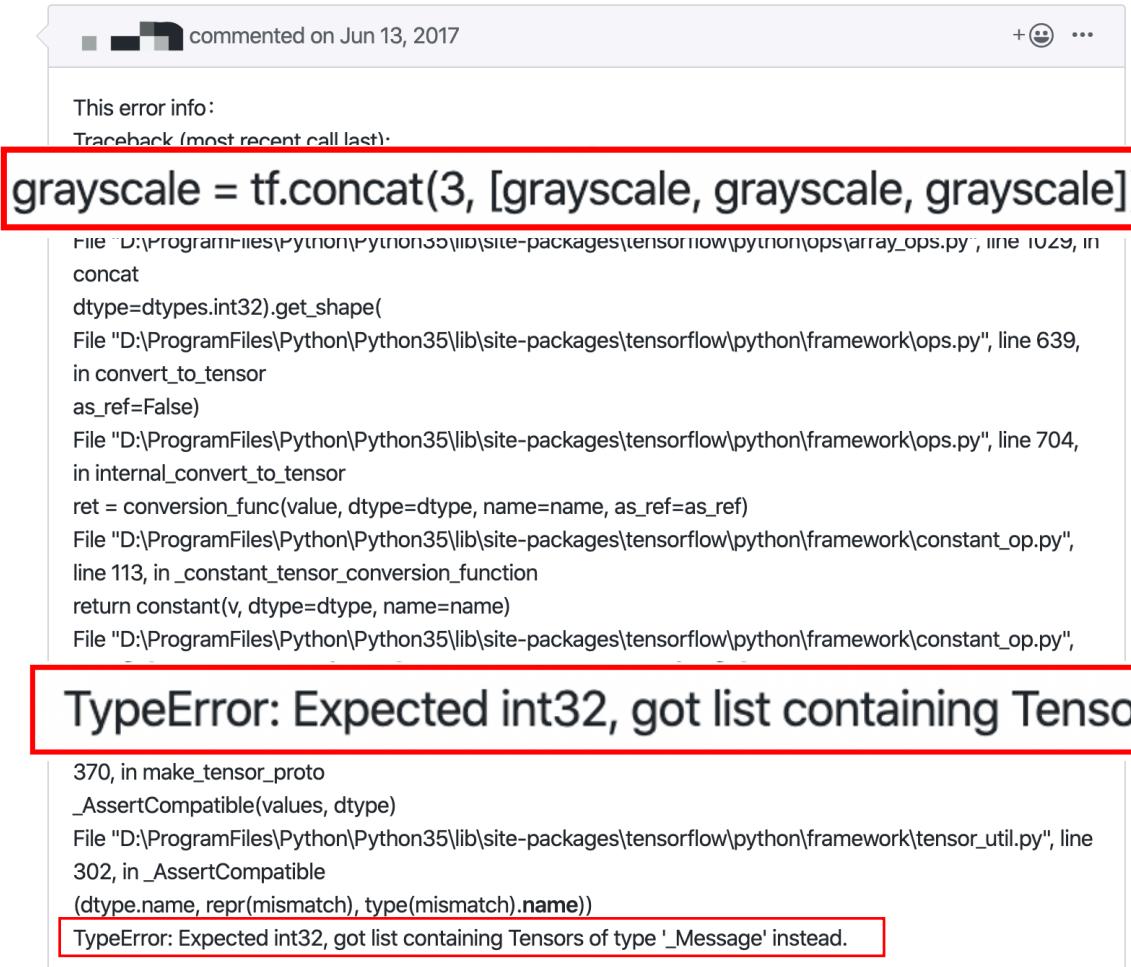
This error info：
Traceback (most recent call last):

grayscale = tf.concat(3, [grayscale, grayscale, grayscale])

File "D:\ProgramFiles\Python\Python35\lib\site-packages\tensorflow\python\ops\array_ops.py", line 1029, in concat
dtype=dtypes.int32).get_shape(
File "D:\ProgramFiles\Python\Python35\lib\site-packages\tensorflow\python\framework\ops.py", line 639, in convert_to_tensor
as_ref=False)
File "D:\ProgramFiles\Python\Python35\lib\site-packages\tensorflow\python\framework\ops.py", line 704, in internal_convert_to_tensor
ret = conversion_func(value, dtype=dtype, name=name, as_ref=as_ref)
File "D:\ProgramFiles\Python\Python35\lib\site-packages\tensorflow\python\framework\constant_op.py", line 113, in _constant_tensor_conversion_function
return constant(v, dtype=dtype, name=name)
File "D:\ProgramFiles\Python\Python35\lib\site-packages\tensorflow\python\framework\constant_op.py",

TypeError: Expected int32, got list containing Tensors of type '_Message' instead.

370, in make_tensor_proto
_AssertCompatible(values, dtype)
File "D:\ProgramFiles\Python\Python35\lib\site-packages\tensorflow\python\framework\tensor_util.py", line 302, in _AssertCompatible
(dtype.name, repr(mismatch), type(mismatch).**name**))
TypeError: Expected int32, got list containing Tensors of type '_Message' instead.

API Evolution Process

Parameter Reordering

```
# TF 0.12
concat(concat_dim, values, name='concat')
```

Parameter Renaming

```
# TF 1.0
concat(values, axis, name='concat')
```

```
tf.concat(3, TypeError (31.4%) scale, grayscale])
```

Fixing Strategy

Altering argument passing (29/347)

```
tf.concat([...],3)

tf.concat( axis=[...],  values=3)
```

Positional Argument
Keyword Argument

parameter reordering would break the API calls with positional arguments

```
tf.nn.softmax(q_values, axis=0)
```

Keyword Argument
Positional Argument

parameter renaming would break the API calls with keyword arguments

Fixing Strategy

- ## Catching related exceptions (32/347)

  Developers often catch such exceptions and fallback to alternative APIs

  ```
  try:
      from pandas import plotting
  except ImportError:
      from pandas.tools import plotting
  ```

- ## Checking library versions or API existence (32/347)

  Explicitly check the library versions or API existence before calling a target API

  ```
  if(hasattr(pandas,"plotting")):
      from pandas import plotting
  else:
      from pandas.tools import plotting
  ```

Empirical Findings (Overall)

14 Evolution patterns

- from pandas import plotting

+ try:
+   from pandas import plotting
+ except ImportError:
+   from pandas.tools import plotting

- from pandas import plotting

+ if(hasatr(pandas,"plotting"))
+   from pandas import plotting
+ else
+   from pandas.tools import plotting

Positional arguments -> keyword arguments
- grayscale = tf.concat([...], 0)
+ grayscale = tf.concat(axis=[...], values=0)

Keyword arguments -> positional arguments
- softmax_out = tf.nn.softmax(q_values, axis=0)
+ softmax_out = tf.nn.softmax(q_values, 0)

3 Major exception types

```
File "D:\ProgramFiles\Python\Python35\lib\site-packages\tensorflow\python\framework\constant_op.py",
line 102, in constant
tensor_util.make_tensor_proto(value, dtype=dtype, shape=shape, verify_shape=verify_shape))
File "D:\ProgramFiles\Python\Python35\lib\site-packages\tensorflow\python\framework\tensor_util.py", line
370, in make_tensor_proto
_AssertCompatible(values, dtype)
File "D:\ProgramFiles\Python\Python35\lib\site-packages\tensorflow\python\framework\tensor_util.py", line
302, in _AssertCompatible
(dtype.name, repr(mismatch), type(mismatch).name))
TypeError: Expected int32, got list containing Tensors of type '_Message' instead.
```

```
    self.models_module = import_module(models_module_name)
File "/opt/python/2.7.9/lib/python2.7/importlib/__init__.py", line 37, in import_module
    __import__(name)
File "/home/travis/build/PetrDlouhy/django-polymorphic/polymorphic/models.py", line 31, in <m
    class PolymorphicModel(six.with_metaclass(PolymorphicModelBase, models.Model)):
File "/home/travis/virtualenv/python2.7.9/lib/python2.7/site-packages/django/utils/six.py", l
    return meta(name, bases, d)
File "/home/travis/build/PetrDlouhy/django-polymorphic/polymorphic/base.py", line 90, in __ne
    new_class._default_manager = user_manager._copy_to_model(new_class)
AttributeError: 'PolymorphicManager' object has no attribute '_copy_to_model'
```

4 Fixing strategies

PyCompact

A compatibility issue detector

Workflow

Library Releases

API Knowledge Base Extraction

API Knowledge Base

Python Source File

.PY

Issue Detection

Compatibility Issues
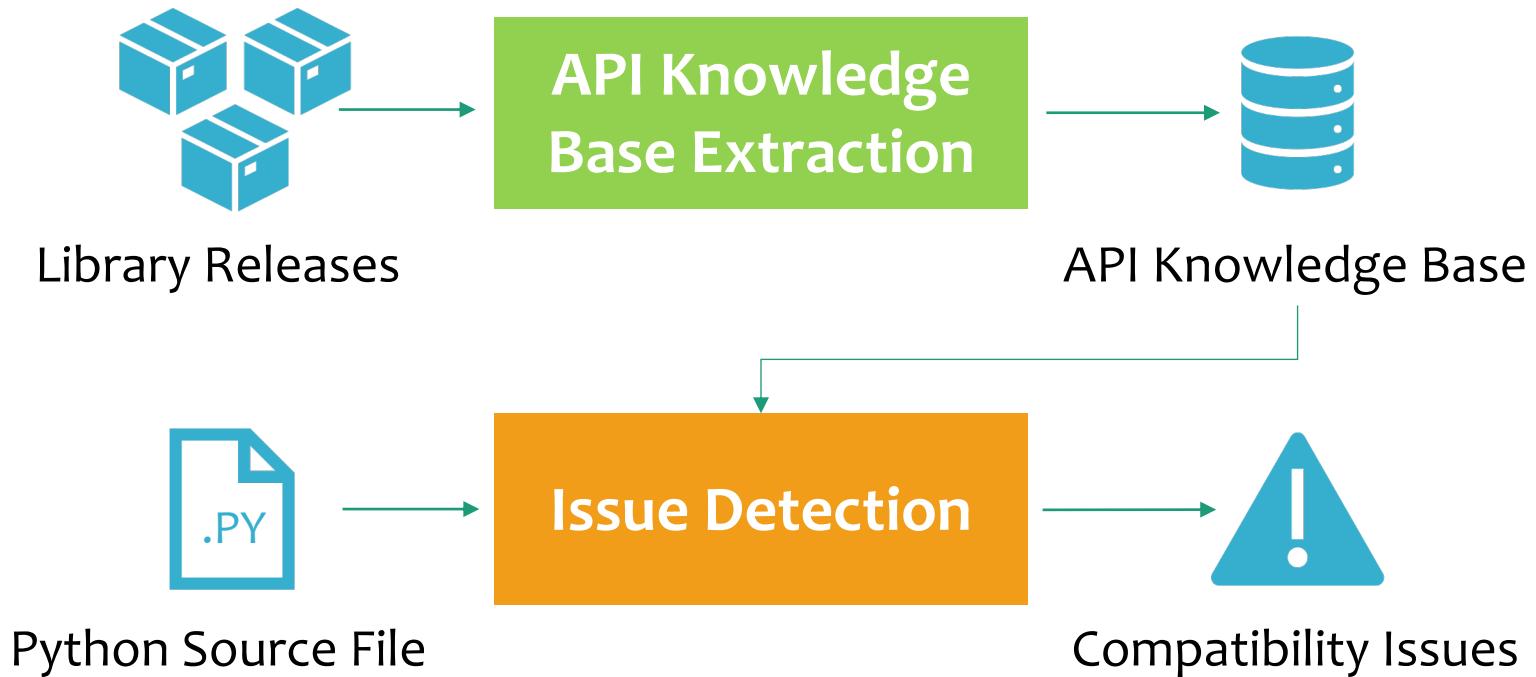
Tool (Demo)

API Knowledge Base

```
tf.pack(values , name="pack")
                    │ tf 0.x to 1.x
                    ▼
tf.stack(values , name="pack")


tf.SparseTensor(..., shape=...)
                    │ tf 0.x to 1.x
                    ▼
tf.SparseTensor(..., dense_shape=...)
```

```
1 ....
2 ....
3 if hasattr(tf, 'stack'):
4     stack = tf.stack
5     unstack = tf.unstack
6 else:
7     stack = tf.pack
8     unstack = tf.unpack
9 ....
10 a = collections.OrderedDict([
11        ((0, 0, 0), 1),
12        ((0, 0, 1), 2),
13 ....
14     ])
15 a = tf.SparseTensor(list(a.keys()),
list(a.values()), dense_shape=[2,2,2])
16 ...
```

Tool (Demo)

API CallSite	Fixed	Warning
(L4) tf.stack	Yes	No
(L5) tf.unstack	Yes	No
(L7) tf.pack	Yes	No
(L8) tf.unpack	Yes	No
(L15) tf.S..T..	No	Yes

ARR: API Rename or
Parameter Rename
Relocation

```
1  ....
2  ....
3  if hasattr(tf, 'stack') \
       and hasattr(tf, 'unstack'):
4      stack = tf.stack
5      unstack = tf.unstack
6  else:
7      stack = tf.pack
8      unstack = tf.unpack
9  ....
10 a = collections.OrderedDict([
11         ((0, 0, 0), 1),
12         ((0, 0, 1), 2),
13 ....
14     ])
15 a = tf.SparseTensor(list(a.keys()),
list(a.values()), dense_shape=[2,2,2])
16 ...
```
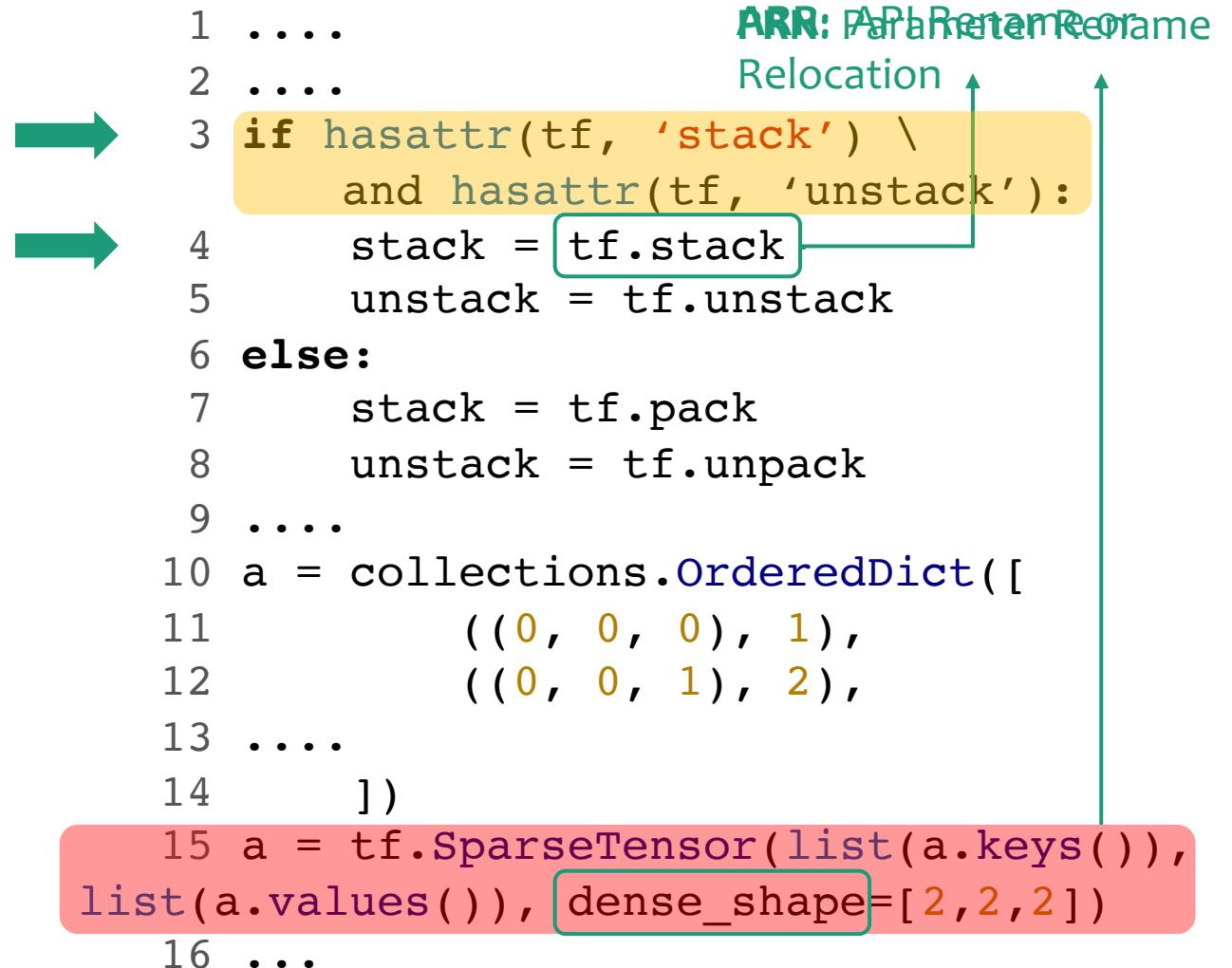
Evaluation of PyCompact

13 Issue reports   **7** Confirmed   **3** Ignored

Repository Name	Revision	# Stars	# API Call Sites	# ARR	# PRN	# TP	# FP	Issue ID(s)
ck-tensorflow [35]	c1ff67d	77	446	3	0	3	0	114 (Fixed)
keras [36]	9d33a02	40,945	612	1	1	2	0	12710 (Ignored)
tensorflow-yolov3 [37]	970cd22	783	340	7	8	15	0	117 (Fixed), 136 (Pending)
tensorflow-101 [38]	aa80965	1,041	2,650	42	21	63	0	19 (Fixed), 21 (Fixed)
tensorflow-playground [39]	349db97	71	272	12	0	12	0	8 (Pending)
TF-and-DL-Tutorial [40]	1b70523	2,234	798	31	0	1	30	38 (Fixed)
Valified_Code_Classify [41]	9043f4a	77	69	1	0	1	0	3 (Pending)
spotify-tensorflow [42]	5f607ad	79	129	0	1	1	0	193 (Confirmed)
R2CNN_FPN_Tensorflow [43]	0dc1b20	317	2,097	59	71	130	0	66 (Ignored), 68 (Ignored)
tensorflow-triplet-loss [44]	fc69836	522	180	0	3	3	0	41 (Confirmed)
Total			7,593	156	105	231	30	13

Future Work

Fully automate PyCompat

Patching of compatibility issues

API evolution of other DPL frameworks

Thanks

Q&A