few-shot

June 29, 2022

[ ]:
```
import spacy
import classy_classification
```

[ ]:
```
import pandas as pd
df = pd.read_csv("test.csv")
data = {}
sample_size = 10

candidate_labels = ["Chemicals", "Construction Materials", "Containers and␣
↪Packaging", "Metals and Mining", "Paper and Forest Products"]

for label in candidate_labels:
    candidate_values = df.query(f"`Level 3` == '{label}'").
↪sample(n=sample_size)['BusinessDesc'].values.tolist()
    data[label] = candidate_values
```

[ ]:
```
nlp = spacy.blank("en")
nlp.add_pipe(
    "text_categorizer",
    config={
        "data": data,
        "model": "sentence-transformers/all-mpnet-base-v2",
        "device": "gpu"
    }
)
```

```
WARNING: The shape inference of prim::Constant type is missing, so it may result
in wrong shape inference for the exported graph. Please consider adding it in
symbolic function.
WARNING: The shape inference of prim::Constant type is missing, so it may result
in wrong shape inference for the exported graph. Please consider adding it in
symbolic function.
WARNING: The shape inference of prim::Constant type is missing, so it may result
in wrong shape inference for the exported graph. Please consider adding it in
symbolic function.

Ignore MatMul due to non constant B: /[MatMul_215]
Ignore MatMul due to non constant B: /[MatMul_221]
Ignore MatMul due to non constant B: /[MatMul_310]
```

```
Ignore MatMul due to non constant B: /[MatMul_316]
Ignore MatMul due to non constant B: /[MatMul_405]
Ignore MatMul due to non constant B: /[MatMul_411]
Ignore MatMul due to non constant B: /[MatMul_500]
Ignore MatMul due to non constant B: /[MatMul_506]
Ignore MatMul due to non constant B: /[MatMul_595]
Ignore MatMul due to non constant B: /[MatMul_601]
Ignore MatMul due to non constant B: /[MatMul_690]
Ignore MatMul due to non constant B: /[MatMul_696]
Ignore MatMul due to non constant B: /[MatMul_785]
Ignore MatMul due to non constant B: /[MatMul_791]
Ignore MatMul due to non constant B: /[MatMul_880]
Ignore MatMul due to non constant B: /[MatMul_886]
Ignore MatMul due to non constant B: /[MatMul_975]
Ignore MatMul due to non constant B: /[MatMul_981]
Ignore MatMul due to non constant B: /[MatMul_1070]
Ignore MatMul due to non constant B: /[MatMul_1076]
Ignore MatMul due to non constant B: /[MatMul_1165]
Ignore MatMul due to non constant B: /[MatMul_1171]
Ignore MatMul due to non constant B: /[MatMul_1260]
Ignore MatMul due to non constant B: /[MatMul_1266]
```

[ ]:
```
<classy_classification.classifiers.spacy_few_shot_external.classySpacyFewShotExt
ernal at 0x7f837ffb6fa0>
```

[ ]:
```python
import csv

num_rows = 0
correct_rows = 0

reader = csv.DictReader(open('test.csv'))

with open('output_file.csv', 'w', newline='') as csv_file:
    fieldnames = ['Level 3', 'Prediction', 'BusinessDesc', 'split']
    writer = csv.DictWriter(csv_file, fieldnames, delimiter='\t')
    writer.writeheader()
    for row in reader:
        num_rows += 1
        doc = row['BusinessDesc']
        categories = nlp(doc)._.cats
        max_category = max(categories, key=categories.get)
        write_row = row.copy()
        write_row['Prediction'] = max_category
        writer.writerow(write_row)
        if write_row['Level 3'] == write_row['Prediction']:
            correct_rows += 1
```

```
print("Accuracy: "+ str(correct_rows/num_rows*100) + "%")
```

```
Accuracy: 83.7887067395264%
```