Crowd Sourcing Training Data For Machine Learning

with Amazon Mechanical Turk

Prerequisites

- Python 2.7 or 3.5
- Amazon AWS account
- MTurk Requester Account
- AWS Command Line Tool
- IAM User Profile with
  - AmazonMechanicalTurkCrowdFullAccess
  - AmazonMechanicalTurkFullAccess

Natural Language Processing APIs Available

- sentiment-analysis
- emotion-detection
- named-entity-recognition
- coreference-resolution
- key-phrase-extraction
- semantic-similarity
- collect-utterance-for-intent
- intent-detection
- text-categorization

Computer Vision APIs Available

- bounding-box
- image-contains
- image-categorization
- image-similarity

Install Python client and Boto 3

```
pip install --upgrade mturk-crowd-beta-client

    --ignore-installed six
```

Build Python code

```
from mturk_crowd_beta_client import MTurkCrowdClient
from boto3.session import Session
session = Session(profile_name='mturk-crowd-caller')
crowd_client = MTurkCrowdClient(session)

api_to_use = 'sentiment-analysis'
task_name = 'pycascades2018'
text = {'text':'Off to a fantastic start at @pycascades.'}
put_result = crowd_client.put_task(api_to_use, task_name, text)
get_result = crowd_client.get_task(api_to_use, task_name)
```

Jupyter Notebook - github/aws-samples/mturk-jupyter

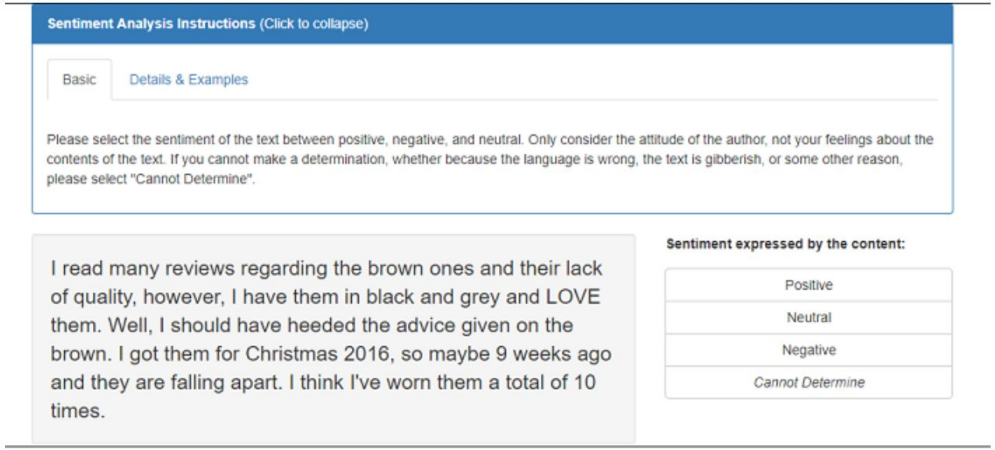
sentiment-analysis

Input: `{ "text": "Everything is wonderful!" }`

Result: `{'sentiment': 'positive'}`

Max length of the input text is 400 characters. Sentiment is one of positive, negative, neutral or cannot determine.

When you create a Task using the sentiment-analysis API, you're automatically creating a Human Intelligence Task (HIT) on worker.mturk.com. Here's an example of a sentiment analysis HIT.



For more information, refer to the sentiment-analysis API documentation.

Jupyter Notebook - github/aws-samples/mturk-jupyter

Create a Task

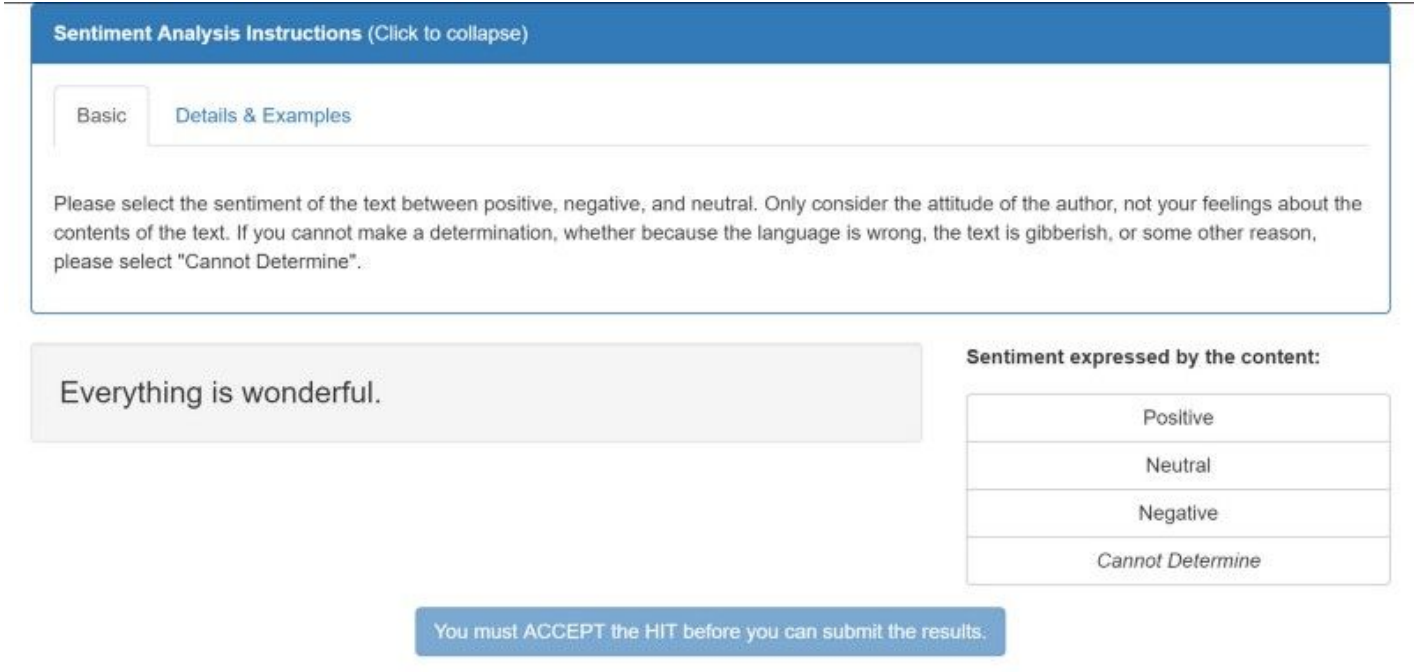
In [ ]:
```
#set the function_name to the name of the API
function_name = 'sentiment-analysis'
```

In [ ]:
```
# automatically generate a random task ID
task_name = 'my-test-task-' + uuid.uuid4().hex
print(task_name)
```

In [ ]:
```
# define the text that you want analyzed, up to 400 characters
text = 'The trip by @VP Pence was long planned. He is receiving great praise for leaving game after the pl
ayers showed such disrespect for country!'
```

In [ ]:
```
# create a single task with the input you specified above
put_result = crowd_client.put_task(function_name,
                                   task_name,
                                   {'text': text})
print('PUT response: {}'.format(
    {'status_code': put_result.status_code, 'task': put_result.json()}))
```

Worker Preview



Jupyter Notebook - github/aws-samples/mturk-jupyter

Get the result

Wait a few minutes before calling get_task to give Workers a chance to submit answers.

In [ ]:
```
get_result = crowd_client.get_task(function_name, task_name)

print('GET response: {}'.format(
    {'status_code': get_result.status_code, 'task': get_result.json()}))
```

Sample Response

Example response for a successful Task:

```
{
  "taskName": "my-task-name",
  "input": { "text": "Python is wonderful!"   },
  "problemDetails": null,
  "state": "completed",
  "result": { "sentiment": "positive"  }
}
```