問題1 配点45点

画像の手書き文字を判別するプログラムを作成します。以下の提供ファイルを使用し、要求事項を全て満たすようにプログラムしてください。

プログラムは「kouka\_1\_1.ipynb」を修正して作成してください。

作成したプログラムはファイル名を「kouka\_1\_1\_名前.ipynb」に変更して提出してください。

提供ファイル群

1. kouka1\_1.ipynb　…　このプログラムの空欄を実装してください。
2. kouka\_1\_1\_data.zip　…　以下のファイルが圧縮されています。

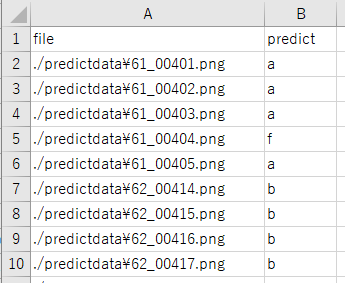
|  |  |
| --- | --- |
| traindata (フォルダ) | 手書きのアルファベット1文字（a～g）の画像データが入っています |
| predictdata (フォルダ) | 手書きのアルファベット1文字（a～g）の画像データが入っています |
| index\_train.csv | traindata内のファイル名と画像に書かれている文字(正解)の一覧 |
| sample\_model.pkl | 手書きアルファベットを判別するモデルのサンプル |

※画像ファイルは32ピクセルx32ピクセル

要求事項

1. traindata内の画像ファイルと「index\_train.csv」を使用して、手書き文字を認識するモデルを作成する。作成したモデルは「kouka\_1\_model\_名前.pkl」として保存する。
2. 保存された学習済みモデルを読み込み、predictdata内のすべての画像ファイルに対して、画像に書かれている文字を判別する。  
   判別結果は「file, predict」のCSV形式で、「result.csv」として保存する。  
   「file」列は判別対象の画像ファイルのパス、「predict」列は判別結果の文字を記載する。
   * (1)で作成した学習済みモデル、もしくは提供された「sample\_model.pkl」のどちらを使用しても構いません

図1 「result.csv」の抜粋



問題2 配点25点

画像ファイルの四角枠の中に、対象の数値が含まれているかどうかをチェックするプログラムを作成します。以下の提供ファイルを使用し、要求事項を全て満たすようにプログラムしてください。

プログラムは「kouka\_1\_2.ipynb」を修正して作成してください。

作成したプログラムはファイル名を「kouka\_1\_2\_名前.ipynb」に変更して提出してください。

提供ファイル群

1. kouka1\_2.ipynb　…　このプログラムを、コメントを参考に実装してください。
2. kouka\_1\_2\_data.zip　…　以下のファイルが圧縮されています。

|  |  |
| --- | --- |
| digits.pkl | 8ピクセル×8ピクセルの画像から数字を判別するモデル。 |
| target.png | 数値の書かれた画像ファイル。  ここに書かれている数値を探す対象とする |
| check1.png～check5.png | 数値の書かれた画像ファイル。  探す対象の数値が含まれているかチェックする |

要求事項

1. 図2「読み込む画像の例」のように、四角枠のある画像ファイルを読み込み、四角枠の領域を抽出する関数を作成する。
2. 上記で作成した関数を使用し、画像ファイルの四角枠の領域に書かれている数値を判別する関数を作成する。数値の判別には「digits.pkl」を使用する。
3. 「target.png」に書かれた数値を判別し、表示する。（以下、この数値を「ターゲット数値」と呼ぶ）
4. 「check1.png」～「check5.png」の5つの画像を全てチェックし、ターゲット数値が含まれているか判別する。  
   含まれている場合は、チェックした画像のファイル名と、その画像に含まれていたターゲット数値を出力する。  
   含まれていない場合は、チェックした画像のファイル名と「含まれない」を表示する。

図2 読み込む画像の例



問題3 配点30点

マルコフ辞書を使用し、夏目漱石風の文章を自動作成するプログラムを作成します。  
以下の提供ファイルを使用し、要求事項を全て満たすようにプログラムしてください。

プログラムは「kouka\_1\_3.ipynb」を修正して作成してください。

提供ファイル群

1. kouka1\_3.ipynb　…　このプログラムの空欄を実装してください。
2. kouka\_1\_3\_data.txt　…　マルコフ辞書を作成する際の元にするテキストファイル。  
   (青空文庫より 夏目漱石　「吾輩は猫である」の抜粋)

要求事項

以下の処理を盛り込むこと

1. テキストファイルを1行ずつ読み込み、形態素解析する。
2. 形態素解析で得られた単語をマルコフ連鎖辞書に登録する。
3. マルコフ連鎖辞書に登録する単語で、「名詞」のものと出現回数の一覧を表示する。
4. 作成したマルコフ連鎖辞書を使用して、Botに3つ以上の文章を作成させる。