THREE PHILOSOPHERS WALK INTO A CLASSIFIER

Author: J. Horner, jhorner@cybermesa.com
Last modified: 8 March 2021/1045 CT

Platform:
  -- *Mathematica* (Wolfram Research 2021)
  -- Windows 10
  -- Dell Inspiron 545 desktop with
        -- Intel Q8200 quadprocessor clocked at 2.33 GHz
        -- 8 GB RAM
        -- 1 TB disk
  -- 1 Mb/s Internet access

1.0 Introduction

This Mathematica notebook illustrates some of *Mathematica*'s machine-learning functionality. The notebook creates and trains a classifier based on two works of each of Plato, Kant, and Bertrand Russell. Given a third work, T, of each author and not contained in the training set of any author, the classifier identifies the author of T.

The implementation in Section 2.0 was adapted from an example in the *Mathematica* online documentation (Wolfram Research 2021) for entry "Classify" .

2.0 Mathematica code

Import three samples (from Project Gutenberg) from three authors.   The first two samples in the list for each author are used to train the classifier; the third sample is used to test the performance of the classifier.

Plato:

```
In[1]:= Republic = Import["http://www.gutenberg.org/cache/epub/1497/pg1497.txt"];
        Meno = Import["http://www.gutenberg.org/cache/epub/1643/pg1643.txt"];
        Phaedo = Import["http://www.gutenberg.org/cache/epub/1658/pg1658.txt"];
```

Kant:

```
In[4]:= CritiquePureReason = Import["http://www.gutenberg.org/files/4280/4280-0.txt"];
        MetaphysicMorals = Import["http://www.gutenberg.org/cache/epub/5682/pg5682.txt"];
        ProlegomenaMetaphysics = Import["http://www.gutenberg.org/files/52821/52821-0.txt"];
```

Bertrand Russell:

```
In[7]:= ProblemsPhilosophy = Import["http://www.gutenberg.org/cache/epub/5827/pg5827.txt"];
        AnalysisMind = Import["http://www.gutenberg.org/cache/epub/2529/pg2529.txt"];
        KnowledgeExternalWorld = Import["http://www.gutenberg.org/files/37090/37090-0.txt"];
```

Train a classifier, **author**, using the first two examples for each author above. Let Mathematica choose the classification method.

```
In[10]:= author =
         Classify[<|"Plato" → {Republic, Meno}, "Kant" → {CritiquePureReason, MetaphysicMorals},
           "Russell" → {ProblemsPhilosophy, AnalysisMind}|>]
```

```
Out[10]= ClassifierFunction[ ⊞  ⣿  Input type: Text
                                    Classes: Kant, Plato, Russell  ]

                             ⤵  Data not in notebook; Store now »
```

Show some information about the classifier.

```
In[11]:= Information[author]
```

Out[11]=

Classifier information	
Data type	Text
Classes	Kant, Plato, Russell
Accuracy	(67. ±44.)%
Method	Markov
Single evaluation time	924. ms/example
Batch evaluation speed	1.08 examples/s
Loss	0.824 ± 0.32
Model memory	1.06 MB
Training examples used	6 examples
Training time	29.5 s

Mathematica selected the Hidden Markov Model method (see Petrushin 2000) for the classifier.

3.0 Results

Ask the classifier to identify the author of the third example in each of the author-lists above.

In[12]:= `author[{Phaedo, ProlegomenaMetaphysics, KnowledgeExternalWorld}]`

Out[12]= `{Plato, Kant, Russell}`

The total time to execute the code (including accessing and downloading the entire works indicated from the Project Gutenberg site) in Sections 2 and 3 on the platform identified above was about 1 minute.

4.0 Discussion

The example in Sections 2 and 3 correctly identifies the authors of the works in the test set, based on just two training samples per author, which is surprisingly good behavior. In general, machine-learning performance is sensitive to the content of training and test sets, and to the learning algorithm selected (Russell and Norvig 2016).

In principle, the example in Sections 2 and 3 can be used as a template for any digitized training and test items. For example, one could use the same rubric to classify music scores encoded in MusicXML (W3C Community Group 2017), sound spectra, and images, among others.

5.0 References

Petrushin VA. (2000). *Hidden Markov Models: Fundamentals and Applications*. https://www.eecis.udel.edu/~lliao/cis841s06/hmmtutorialpart1.pdf and https://www.eecis.udel.edu/~lliao/cis841s06/hmmtutorialpart2.pdf. Accessed 8 March 2021.

Project Gutenberg. http://www.gutenberg.org/. Accessed 8 March 2021.

Russell S and Norvig P. (2016). *Artificial Intelligence: A Modern Approach.* 3rd edition. Pearson.

Wolfram Research. (2021). *Mathematica Home Edition* v12.1. https://www.wolfram.com/mathematica-home-edition/. Accessed 8 March 2021.

W3C Community Group. (2017). *MusicXML. V3.1.* Final Community Group Report. https://www.w3.org/2017/12/musicxml31/. Accessed 8 March 2021.