Package ‘rethnicity’

November 30, 2021

<table>
<thead>
<tr>
<th>Type</th>
<th>Package</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Predicting Ethnic Group from Names</td>
</tr>
<tr>
<td>Version</td>
<td>0.2.2</td>
</tr>
<tr>
<td>Maintainer</td>
<td>Fangzhou Xie <a href="mailto:fangzhou.xie@rutgers.edu">fangzhou.xie@rutgers.edu</a></td>
</tr>
<tr>
<td>Description</td>
<td>Implementation of the ethnicity prediction method, described in &quot;Rethnicity: Predicting Ethnicity from Names&quot; by Fangzhou Xie (2021) <a href="">arXiv:2109.09228</a>.</td>
</tr>
<tr>
<td>License</td>
<td>CC BY-NC-SA 4.0</td>
</tr>
<tr>
<td>Encoding</td>
<td>UTF-8</td>
</tr>
<tr>
<td>RoxygenNote</td>
<td>7.1.2</td>
</tr>
<tr>
<td>URL</td>
<td><a href="https://github.com/fangzhou-xie/rethnicity">https://github.com/fangzhou-xie/rethnicity</a></td>
</tr>
<tr>
<td>BugReports</td>
<td><a href="https://github.com/fangzhou-xie/rethnicity/issues">https://github.com/fangzhou-xie/rethnicity/issues</a></td>
</tr>
<tr>
<td>Depends</td>
<td>R (&gt;= 3.4.0)</td>
</tr>
<tr>
<td>LinkingTo</td>
<td>Rcpp, RcppEigen, RcppThread</td>
</tr>
<tr>
<td>Imports</td>
<td>Rcpp, cli</td>
</tr>
<tr>
<td>Suggests</td>
<td>knitr, rmarkdown, testthat (&gt;= 3.0.0), magrittr, parallel</td>
</tr>
<tr>
<td>VignetteBuilder</td>
<td>knitr</td>
</tr>
<tr>
<td>Config/testthat/edition</td>
<td>3</td>
</tr>
<tr>
<td>NeedsCompilation</td>
<td>yes</td>
</tr>
<tr>
<td>Author</td>
<td>Fangzhou Xie [aut, cre] (<a href="https://orcid.org/0000-0001-7702-093X">https://orcid.org/0000-0001-7702-093X</a>)</td>
</tr>
<tr>
<td>Repository</td>
<td>CRAN</td>
</tr>
<tr>
<td>Date/Publication</td>
<td>2021-11-30 09:30:02 UTC</td>
</tr>
</tbody>
</table>

R topics documented:

- predict_ethnicity ...................................................... 2
- predict_fullname .......................................................... 3
- predict_lastname .......................................................... 3

Index 5
predict_ethnicity  

Predict ethnicity from names.

Description

Predict ethnicity either by last names or both first and last names. This is the default and recommended method for prediction.

Usage

predict_ethnicity(
  firstnames = NULL,
  lastnames = NULL,
  method = "fullname",
  threads = 0,
  na.rm = FALSE
)

Arguments

- **firstnames**: A character vector of first names. Default to NULL. Only use this if you are using 'method' = 'fullname'.
- **lastnames**: A character vector of last names. Default to NULL. Use this in both 'fullname' and 'lastname' methods.
- **method**: "fullname" or "lastname". Inference method to choose from.
- **threads**: single integer. Number of threads to use for multi-threading.
- **na.rm**: TRUE or FALSE (bool). If TRUE, then the NAs will be removed; if FALSE, then return error if there is NA in the arguments.

Value

data.frame with probability of being each ethnic group and the predicted group (one with highest probability)

Examples

predict_ethnicity(firstnames = "Alan", lastnames = "Turing")
**predict_fullname**

*Predict ethnicity from full name*

**Description**

Predicts ethnicity from first names and last names, using self-trained model with customized labels. This is designed for advanced users who wish to use their own models. For most use cases, use `predict_ethnicity()` for prediction.

**Usage**

```r
predict_fullname(
  firstnames,
  lastnames,
  na.rm = FALSE,
  threads = 0L,
  labels = NULL,
  model_path = NULL
)
```

**Arguments**

- `firstnames` character vector, first names
- `lastnames` character vector, last names
- `na.rm` bool, default to FALSE, whether to remove the na in the ‘lastnames’
- `threads` int, number of threads for multi-threading
- `labels` character vector, labels of the classification model, needs to be in the same order as the trained model
- `model_path` character file path, the path to the trained model in .json format (converted from Keras by frugally-deep)

**Value**

data.frame with predicted probability and predicted ethnicity

---

**predict_lastname**

*Predict ethnicity from last name*

**Description**

Predicts ethnicity from last names, using self-trained model with customized labels. This is designed for advanced users who wish to use their own models. For most use cases, use `predict_ethnicity()` for prediction.
Usage

predict_lastname(
  lastnames,
  na.rm = FALSE,
  threads = 0L,
  labels = NULL,
  model_path = NULL
)

Arguments

lastnames  character vector, last names
na.rm      bool, default to FALSE, whether to remove the na in the 'lastnames'
threads    int, number of threads for multi-threading
labels     character vector, labels of the classification model, needs to be in the same order
            as the trained model
model_path character file path, the path to the trained model in .json format (converted from
            Keras by frugally-deep)

Value

data.frame with predicted probability and predicted ethnicity
Index

predict_ethnicity, 2
predict_fullname, 3
predict_lastname, 3