虚假信息检测数据集说明

数据集包含训练数据、测试数据、事实审查数据,均为csv为格式。数据集中的数据标签根据虚假程度分为6个类别，分别为极度虚假(pants-fire)、虚假(false)、大部分虚假(barely-true)、半真半假(half-true)、大部分真实(mostly-true)、真实(true)。

（1）训练集包含一批网络文本信息样本及其标签，文件名为train.csv，格式如下：

|  |  |
| --- | --- |
| train.csv | |
| ID | 样本ID |
| author | 发布人名称 |
| claim | 网络文本信息 |
| label | 类别标签 |

（2）测试集包含一批不含标签的样本，文件名为test.csv，格式如下：

|  |  |
| --- | --- |
| test.csv | |
| ID | 样本ID |
| author | 发布人名称 |
| claim | 网络文本信息 |

（3）事实审查数据包含20000条人工编写的证据文本，可供灵活使用，文件名为evidence.csv，格式如下：

|  |  |
| --- | --- |
| evidence.csv | |
| ID | 证据ID |
| evidence | 证据文本 |