# 数据统计

样本总数：6439，包含

零售商样本：2373

生产商样本：1379

其他：2687

# 十折交叉验证结果

训练用搜索结果和网站描述，测试用搜索结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Micro f1 | LR | SVM | CNN | BiLSTM | Ours |
| BOW(1-gram) | 0.681 | 0.645 |  |  |  |
| BOW(2-gram) | 0.653 | 0.427 |  |  |  |
| BOW(3-gram) | 0.580 | 0.417 |  |  |  |
| One-hot | 0.520 | 0.420 | 0.405 | 0.551 | 0.647 |
| BERT | 0.510 | 0.708 | 0.606 | 0.637 | 0.725 |

训练用搜索结果和网站描述，测试用搜索结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Macro f1 | LR | SVM | CNN | BiLSTM | Ours |
| BOW(1-gram) | 0.511 | 0.565 |  |  |  |
| BOW(2-gram) | 0.503 | 0.220 |  |  |  |
| BOW(3-gram) | 0.443 | 0.200 |  |  |  |
| One-hot | 0.380 | 0.200 | 0.240 | 0.507 | 0.620 |
| BERT | 0.437 | 0.643 | 0.507 | 0.556 | 0.701 |

对比测试时使用搜索结果+网站描述和只用搜索结果的结果

|  |  |  |  |
| --- | --- | --- | --- |
| Micro f1 | 搜索结果+网站描述 | 搜索结果 | 下降程度 |
| LR | 0.525 | 0.510 | 0.015 |
| SVM | 0.738 | 0.708 | 0.030 |
| CNN | 0.735 | 0.606 | 0.129 |
| BiLSTM | 0.761 | 0.638 | 0.123 |
| Ours | 0.768 | 0.725 | 0.043 |

对比测试时使用搜索结果+网站描述和只用搜索结果的结果

|  |  |  |  |
| --- | --- | --- | --- |
| Macro f1 | 搜索结果+网站描述 | 搜索结果 | 下降程度 |
| LR | 0.497 | 0.437 | 0.060 |
| SVM | 0.708 | 0.643 | 0.085 |
| CNN | 0.705 | 0.576 | 0.129 |
| BiLSTM | 0.735 | 0.557 | 0.178 |
| Ours | 0.745 | 0.701 | 0.044 |