### 第二组实验：加入学习排序算法的效果验证

学习排序算法能够利用查询和排序列表的其他特征学习、预测需求跟踪链接，本文使用IR SVM来提升需求跟踪任务的精确度。本节对使用学习排序算法前后的实验结果进行对比，验证使用IR SVM算法对提高软件需求跟踪任务的效果；实验使用MRR和MAP作为评估指标，使用4.5节确定的五个特征作为学习排序模型中构建文档向量的特征值。为保证实验的可靠性，进行十折的交叉验证，最终取交叉中所有验证的平均值为结果。

图20展示使用IR SVM前后的MRR和MAP指标的值。与执行IR SVM之前的结果相比，执行算法之后，各个数据集上的MRR和MAP的值都有所提升，总体来说，MRR的平均值提升了7.48%，MAP的平均值提升了22.05%。MAP的提升，说明使用IR SVM能够提升候选结果列表的排序水平和结果精度；MRR的提升，说明IR SVM算法能使候选结果列表出现正确文档的位置更靠前。也就是说，使用了IR SVM排序后，能够提供更准确的跟踪链接。

图20 执行改进的学习排序算法前后的MRR和MAP比较

本组实验说明，基于计算文档间文本相似度的算法能够取得较好的效果，但只考虑了一个特征是不够的，通过将更多的文本特征与机器学习算法相结合，可以提高需求追踪的效果，找到更多，更准确的追踪链。

### 第三组实验：Tr-WELR模型对比ENRL方法

为便于比较，本组实验采用论文【17】中ENRL使用的数据集eTOUR和EasyClinic，同时使用MRE作为评估的指标。在ENRL中，将12种NLP方法和4种机器学习算法进行组合，对数据进行处理比较，其中对上述两个数据集的最优组合如表9所示。

表9 ENRL实验在eTOUR和EasyClinic上的最优组合

|  |  |  |
| --- | --- | --- |
| **数据集** | **NLP方法** | **机器学习分类算法** |
| eTOUR | stop-stem, TF, LSA100, cosine | IBk |
| EasyClinic | stop-stem, TF, VSM, cosine | IBk |

stop-stem是指去除停用词与抽取单词词干的方法，TF是指统计词频的方法，LSA100指使用概念数为100的潜在语义方法，VSM代表使用向量空间模型，cosine指计算相似度值时使用cosine函数，IBk表示分类时使用k近邻方法。

使用Tr-WELR模型对上述两个数据集处理，计算出MRE作为评估指标，得到的结果与ENRL方法的对比如表10所示。可以看出，在这两个数据集上，Tr-WELR模型的平均相对误差要比ENRL的更小。

表10 Tr-WELR模型和ENRL在eTOUR和EasyClinic上MRE值比较

|  |  |  |
| --- | --- | --- |
| **方法** | **数据集** | **MRE值** |
| ENRL | eTOUR | 0.0300 |
| EasyClinic | 0.0100 |
| Tr-WELR | eTOUR | **0.0208** |
| EasyClinic | **0.0059** |

该实验结果说明，该方法最终的结果比ENRL方法出现的错误更少，即在需求追踪任务中能够取得更好的效果。说明了本文所提出的需求追踪算法与目前较为先进的需求追踪算法，具有更好的性能。

## 运行时间效率

使用r-WELR模型对各个数据集进行追踪时花费在计算上的运行时间如图《》所示，花费在计算上的运行时间包括文本语义相似度计算时间和执行学习排序算法时间。Mean Retrieval Tim表示相应的数据集上恢复一组跟踪链接需要的平均运行时间，R Doc Nums表示在相应数据集中进行一次查询需要检索的文档数量，Mean Doc Length指的是相应数据集的平均文档长度。根据结果数据容易得到，平均运行时间和检索数目、平均文档长度都是正相关的，需要被检索的文档数量越多，文档的平均长度越长，计算需要花费的时间就越长。

从表中可以看到，在不同的数据集上，进行一次查询的时间最短是2.05秒，最长时间为79.25秒，查询时间的长短主要与数据集的大小相关，对于即时性要求不高的需求追踪任务，该时间消耗是可以接受的。

表11 Tr-WELR模型计算部分的运行时间和相关信息

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **数据集** | | **Mean Retrieval Time(sec)** | R Doc Nums | **Mean Doc Length** |
| CM1-NASA | HL→LL | 2.44 | 53 | 339 |
| GANTT | HL→LL | 2.05 | 69 | 107 |
| eTOUR | UC→CC | 49.42 | 116 | 1766 |
| iTrust | UC→CC | 79.25 | 367 | 2867 |
| EasyClinic | UC->ID | 18.09 | 20 | 604 |
| UC->TC | 27.00 | 63 | 604 |
| UC->CC | 29.78 | 30 | 604 |