**模型训练及管理任务规划和内核方法调用说明**

**——胡杨**

1. 功能模块初步设定
2. 文本分类模型的管理功能最初设定：在项目提交的时候，已经提供给用户若干已经训练完成的模型，训练语料也齐备，可以直接使用，初步设定为三个分类算法实现的文本分类模型，在mysql数据表中有所体现。分类模型捆绑包括分类模型在磁盘上的模型文件和存放在mysql数据库中的分类模型属性。用户可以对模型的信息进行查看，对模型采取激活和冻结的操作，还可以对所有激活的模型根据训练语料的变化进行自动的更新操作。
3. 主要功能描述
4. 模型的展示：mysql数据表tbl\_classifierModel中存储了分类模型的元信息，ClassifierModelService【1】接口提供了模型属性表的全部记录查询和查询出激活的模型以及非激活的模型的方法。模型的信息在界面上以列表的方式展示。
5. （可选任务）模型的激活和冻结：如果是非激活的模型，会出现在冻结列表当中，在“操作”一栏中，会有按钮“激活”，点击激活，触发激活函数激活模型，已经激活的模型出现在激活列表当中；在“操作”一栏中，会有按钮“冻结”，点击冻结，触发冻结函数冻结模型。上面任意一个按钮被触发之后，页面刷新，重新载入激活和冻结的模型列表，激活函数：activeClassifierModel【2】，冻结函数：freezeClassifierModel【3】。
6. 单个模型的重新建模：在每条模型记录的最后，设置一栏“刷新”，该栏设置“刷新”按钮，点击刷新，触发刷新函数刷新模型（相当于重新扫描一遍训练语料文件夹重建模型），模型展示页面也刷新一遍。刷新函数：autoRefreshClassiferModel【4】
7. 所有模型自动检测刷新：这个功能是每次自动扫描训练语料文件夹，如果发现文件夹中的文件出现任何变化，就将全部的模型都重建一遍（相当于对每个模型调用操作3），这个可以在界面上增加一个总的按钮，名为“自动刷新模型”以触发此事件，而且用户在每次登入这个界面的时候也需要出发一次“自动刷新模型”的事件。事件触发函数：autoRefreshAllModel【5】
8. 模型切换数据集格式：数据集格式分为数字型和标称型，在arffName字段体现，dataRaw代表标称型数据，dataFilter代表数字型数据。在模型列表相应字段设置按钮“切换数据集”，系统在两者之间自动切换数据集，切换函数：switchDataArffKind【6】
9. 模型参数修改：模型参数各有不同，所以专门为专业人士设置一个字段用于修改参数，参数以json的方式组织（字符串类型），点击“修改按钮”，执行修改该字符串的功能，这个地方不提供内核方法。
10. 内核方法说明及注意事项

【1】：所属包：ims.nlp.entity.service；类：ClassifierModelService；方法说明：listAll()方法用于查询出所有的模型信息，listByActiveState()方法用于查询出所有的激活模型信息，listByActiveStateNeg()方法用于查询出所有的已冻结模型信息。

【2】：所属包：ims.nlp.classifier.weka；类：WekaClassifierModelManagment；方法说明：传入参数：需要激活的模型id，没有返回值。

【3】：所属包：ims.nlp.classifier.weka；类：WekaClassifierModelManagment；方法说明：传入参数：需要冻结的模型id，没有返回值。

【4】：所属包：ims.nlp.classifier.weka；类：WekaClassifierModelManagment；方法说明：传入参数：需要重建的模型id，返回值是一个bool值，如果是true说明训练语料库确实发生了变化，模型已经重建，前台系统提示“模型重建成功”，如果是false说明训练语料库没有发生任何变化，前台系统提示“模型已经是最新的”。

【5】：所属包：ims.nlp.classifier.weka；类：WekaClassifierModelManagment；方法说明：传入参数：无，返回值是一个int值，代表这次全部刷新影响了多少个模型。

【6】所属包：ims.nlp.classifier.weka；类：WekaClassifierModelManagment；方法说明：传入参数：需要切换数据集的模型id，没有返回值。

1. 其他

备注：1.被冻结的模型，除了“激活”事件可以触发之外，其余主要功能都不能使用，界面上最好显示成灰色的状态，直到激活之后，转入激活模型列表。2.对应数据表字段如果不懂，请参照pd图相应数据表。

任务接受人：王秋森