

解答：

句号的问题

成员写的过程中疏忽了句号的问题

我们会重新审查一遍。

用户非语义分析和语义分析

这个在论文里的分析是：

用户的语义分析就是根据用户的发表的文本信息，分析其中的情感，褒义，贬义。

用户的非语义分析是根据用户的微博，评论发表时间，粉丝数，关注数等非文本信息。

这个我们会加上注释的。

为什么是无监督的

想要训练模型，我们必须得到已经确认为水军的样本，在我们找到的论文里，有关机器学习的方法是人工甄别、买水军等等方法，这种方法一方面**耗时耗力**成本高，一方面有**主观倾向**，而没有用到机器学习的一些传统方法是只针对一部分典型水军的特性进行针对性设计相应的算法，比如用针对抱团的方法找到打广告的水军，这种方法不能解决**其他类型**的水军并且**不能对具体某个用户判断是否是水军**。

我们认为根据传统方法可以有针对性且很有效率的找到一类水军，但是不能针对某个具体的用户对其特征进行判断。我们认为各种各样的水军在活跃时间，粉丝数等等一些用户非语义信息是有共性的，所以我们可以用传统的方法找到一些非常典型明显的水军，提取他们的用户非语义信息进行训练，最后得到的模型可以对各种水军，且单个用户都进行甄别的。

因为水军的标记是自动的，没有人为筛选，所以我们认为是无监督的。

关于正常用户的标记

我们标记最有嫌疑的水军，剩下的便是不是最有嫌疑的水军，所以这样的模型只能甄别最有嫌疑的水军，但是我们可以根据我们筛选过程中各个步骤中筛出的样本进行优化，将样本以不同程度标记：有嫌疑的水军，比较有嫌疑的水军，非常有嫌疑的水军，最有嫌疑的水军，或者给不同的筛选出来的样本加上一个嫌疑程度的权值。

这个我们将马上进行添加修改。

最优加权组合模型

这个是我们查找当中学习的一个名词，认为单一的机器模型有局限性，可以根据具体情况将各种机器学习的模型训练结果组合在一起，找到最有加权组合。