2018年10月30日星期二

下午11时31分

在进行了三个基本的实验后，发现自己的假设与CERT数据集并不符合，因此需要重新考虑问题与假设。

自己重新仔细阅读了数据构造的相关论文，发现了其逻辑关系中用户的人格特征与组织结构共同结合影响了用户的人际关系（Relationship），而与被解雇用户的人际关系进一步影响了用户的工作满意度（Job Satisfactory），从而导致用户高倾向的不守时以及跳槽。

我们的目的就是尽量全面、准确地刻画出用户的JS变量，从而事先选定可疑的内部用户。

我们的研究方法是从现实场景假设出发，提出假设模型，然后在数据实验中进行修补。

从本实验开始，关于同一个主题的多个实验以数字进行标识版本，如之前的不同方法与模型分别被标识为【0.1~0.3】，本版本0.4中的多个变体则使用0.4X的形式进一步区分。

实验研究除了验证模型假设，解决问题之外，还要训练自己分析问题、提出假设、验证假设、修改假设的研究能力，这才是“金钥匙”。

1. 我们的假设

依据自己的工作经验分析，既然JS假设受影响于解雇用户的人际关系，那么刻画一个内部用户的人际关系便是首要的问题，摆在我们面前的是三个具体的问题：

1. 给定用户的LDAP、Psychometric以及Email通信统计信息，如何初步建模用户的人际关系？
2. 初步建模出用户的人际关系（即认识、接触的用户范围后），如何进一步明确不同人际联系的重要性？
3. 依据上述人际关系范围与联结重要性的研究分析，如何进一步刻画解雇用户对个体的JS影响？

为了回答上述问题，我们试着进行做出假设，并通过实验验证。

* 假设一
* 与解雇用户的人际关系影响着个体的JS；
* 人际关系（Relationship, RS）可以从两个方面刻画：

1. 反映在组织结构（Organizational Structure, OS），决定了用户日常工作所自然接触的人际范围；我们自然地分析上下结构（领导上下级关系）以及横向关系（同一个最小组织单位内以及第一个父单位下的兄弟组织人员范围），其中横向关系对于CERT5.2而言就是Team以及同一个Department下的不同Teams；
2. 反映在通讯关系（Email Relationship, ER），决定了用户日常交流的人际范围与频度，其中可以使用0.2实验中的邮件分析方法，从信息量的7元特征来进行刻画频度，并将邮件联系人补充到人际关系图中；

* 最终得到了用户的人际关系图（Relationship Graph, RG），该图中节点标识用户ID以及对应的Big5人格特质；边上存在使用四位编码标识OS位置距离的无向边，以及一条标识邮件通讯信息量的有向边；
* 一般而言，计算RG中单独一条联结的重要性，可以依据下述建议的公式：



* 计算用户的JS\_Risk则可以将解雇用户所在边的重要性相加，作为该用户的JS\_Risk