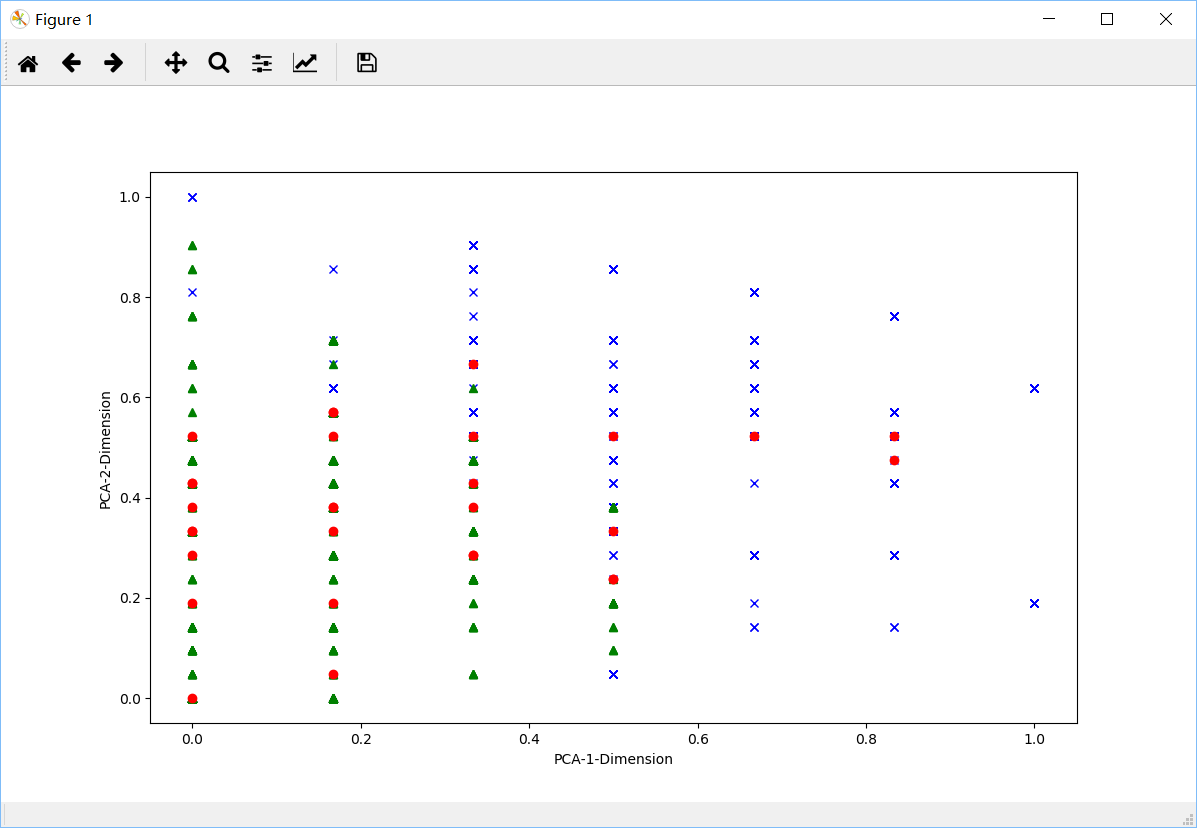
2018年9月28日星期五

下午5时24分

已经下班，自动加班。

随着数据处理开始，自己也着手继续下一步分析。昨天基于CERT5.2中用户离职/解雇的关系数据进行了初步的KMeans聚类分析，但是结果表明无论是PCA还是MinMax都无法挽救该阶段的KMeans，结果如图：



上图为PCA=2时两个维度的图

也许自己不死心吧，决定再次尝试一次，因为论文说“relationship to laid off employees”影响着Job Satisfaction，那么如果直接依据离职人数进行排队呢，不知道场景二的30个用户能排在哪个位置？

首先分析X\_0，即仅考虑同一团队的离职人数，则有：

开始输出Insiders\_2在排序后的索引...

如果考虑X\_1即同一部门内离职人数，依旧不明显

如果考虑LaidOff，依旧结果不明显，此方法作罢。

首先输出X\_0...

30

BYO1846 775

CHP1711 1042

CIF1430 1404

CKP0630 1648

DCC1119 349

GWG0497 1458

HIS1394 463

HMS1658 769

HSN0675 1452

HXP0976 1986

ICB1354 1034

ITA0159 570

JAL0811 1138

KSS1005 1825

LVF1626 105

MCP0611 1555

MDS0680 1654

MGB1235 1964

NAH1366 1557

OKM1092 324

OSS1463 1063

RRS0056 1702

SIS0042 1383

SNK1280 1459

TMT0851 480

TNB1616 447

TRC1838 1155

VCF1602 557

WDT1634 1577

ZIE0741 1830

......<<<<<<分割线>>>>>>......

2018年9月29日星期六

上午10时42分

国庆假期调休。

昨天下班时进一步考证了单独从离职人员数量角度无法有效区分，那么看来需要考虑离职人员的亲密度了。

再继续进行下一步前，简单回顾下自己分析离职人员关系的步骤。

1. 统计了LDAP中所有的部门架构层级关系，进而统计了每个部门架构（细化到团队）中的用户成员，得到了CERT5.2-LDAPUsers.csv
2. 在上述基础上，从2009-12的LDAP文件开始，逐个分析到2011--05的LDAP，统计列出这段时间所有离职（工作时间未干全）的员工列表，以及其离职时间（未离职就是NO），得到了CERT5.2-Leave-Users.csv
3. 在上述基础上，重点统计2000个CERT5.2用户的各个层级架构中，出现的离职时间不晚于其离职的用户成员，并分别列入各个层次的列表中，得到CERT5.2-Leave-Relationship.csv；当过滤掉场景二的攻击者后，得到的是CERT5.2-LaidOff-Relationship.csv

一个简单的示例如：

