2018年10月25日星期四

下午11时17分

实验记录：

今天重新改写了0.1版本的实验程序，将其中聚类的代码部分改写为了计算RelationshipLevel值的形式，然后针对2000个用户计算离职用户与邮件联系人的交叉匹配结果。

第一次实验的结果排序十分不理想：

Insider\_2: 22 : LVF1626 : 8.5834351709

Insider\_2: 269 : TMT0851 : 5.97386395252

Insider\_2: 654 : HIS1394 : 4.42147050456

Insider\_2: 700 : DCC1119 : 4.33994283548

Insider\_2: 706 : TNB1616 : 4.32819849632

Insider\_2: 952 : OKM1092 : 3.80557278275

Insider\_2: 1094 : ITA0159 : 3.54606875786

Insider\_2: 1252 : BYO1846 : 3.20087504496

Insider\_2: 1304 : HSN0675 : 3.08794915103

Insider\_2: 1479 : HMS1658 : 2.69905980496

Insider\_2: 1496 : GWG0497 : 2.65534256707

Insider\_2: 1543 : KSS1005 : 2.54911846516

Insider\_2: 1548 : JAL0811 : 2.53642282297

Insider\_2: 1611 : SNK1280 : 2.37246826979

Insider\_2: 1620 : CHP1711 : 2.35740214145

Insider\_2: 1687 : VCF1602 : 2.13489915235

Insider\_2: 1787 : TRC1838 : 1.79331055601

Insider\_2: 1832 : RRS0056 : 1.51757551633

Insider\_2: 1849 : ICB1354 : 1.45937586266

Insider\_2: 1890 : MGB1235 : 1.13797245136

Insider\_2: 1894 : CIF1430 : 1.08981239594

Insider\_2: 1895 : HXP0976 : 1.07951987208

Insider\_2: 1902 : OSS1463 : 1.04357673488

Insider\_2: 1904 : MDS0680 : 1.04312885453

Insider\_2: 1907 : SIS0042 : 1.00532266726

Insider\_2: 1918 : ZIE0741 : 0.823083125359

Insider\_2: 1919 : MCP0611 : 0.807553699587

Insider\_2: 1920 : NAH1366 : 0.804531198009

Insider\_2: 1921 : WDT1634 : 0.791127853687

Insider\_2: 1966 : CKP0630 : 0.0949496601764

分析原因，排在第一位的用户有很多是单向邮件，而这些邮件无疑影响了我们的判断，因此需要重新排除掉这些单向邮件，只关注ER在(0,1)之间的有效邮件。

2018年10月26日星期五

下午3时35分

鉴于用户关联的用户中有许多是单向邮件，因此决定首先筛查一遍30个关键用户的离职邮件联系人中有多少是非单向的？会不会造成遗漏呢？

实验1：

CERT5.2场景二30个用户的离职邮件联系人分

**BYO1846**

Cluster0

RMB1821,0.28571428571449997,0.0425531914893617,0.005187972305548536,0.0,0.0819672131147541,0.016017337873492615,0.075,0.595974689854,

FDS1841,0.3333333333335,0.0851063829787234,0.01938484458272191,0.027777777777777776,0.13114754098360656,0.053807737409929106,0.109375,0.626952130714,

TRC1838,0.3650793650795,0.48936170212765956,0.033426060894124456,0.033816425120777774,0.6557377049180328,0.003353036762244741,0.0,0.646521851626,

Cluster1

Cluster2

Cluster3

Cluster4

**CHP1711**

Cluster0

KDP1706,0.48387096774195,0.29411764705882354,0.07001791257090592,0.06666666666666667,0.26666666666666666,0.07536764211530689,0.09821428571428571,0.727785393251,

SCO1719,0.3396226415095,0.3529411764705882,0.04587120476849566,0.037037037037,0.5833333333333334,0.07650074786760817,0.0857142857142857,0.680927551366,

Cluster1

Cluster2

Cluster3

Cluster4

**CIF1430**

Cluster0

PTM1432,0.45454545454545,0.5,0.30792675034339395,0.25,0.3529411764705882,0.007346818510967242,0.0,0.805189058915,

Cluster1

Cluster2

Cluster3

Cluster4

**CKP0630**

Cluster0

Cluster1

Cluster2

Cluster3

Cluster4

**DCC1119**

Cluster0

OCW1127,0.4166666666665,0.03164556962025317,0.1347238907365629,0.12,0.030837004405286347,0.04968300754440603,0.0714285714285,0.658472924671,

ELM1123,0.4,0.06329113924050633,0.10997642642577007,0.06,0.06607929515418502,0.06973096286990606,0.06666666666675,0.654722559484,

ACA1126,0.2,0.06962025316455696,0.06875173025643615,0.0727272727272,0.19383259911894274,0.10170705831395398,0.14772727272725,0.587240143021,

KJG1121,0.212121212121,0.04430379746835443,0.13204040116862456,0.17142857142860002,0.11453744493392072,0.1410681379528523,0.1634615384615,0.586773572388,

Cluster1

Cluster2

Cluster3

Cluster4

GKW0043,0.4513888888889,0.4113924050632911,0.07882684877146787,0.0861538461538,0.3480176211453745,0.05327507643453389,0.08544303797475,0.726920220237,

**GWG0497**

Cluster0

Cluster1

Cluster2

Cluster3

Cluster4

**HIS1394**

Cluster0

LSM1382,0.4871794871795,0.48717948717948717,0.0365403969438059,0.02923976608188889,0.25316455696202533,0.2095663757491733,0.15000000000000002,0.746847900653,

KBC1390,0.848484848485,0.717948717948718,0.07388213339402229,0.0793650793651111,0.06329113924050633,0.07836522913657452,0.04000000000000001,0.600632534701,

IVS1411,0.5090909090909,0.717948717948718,0.08785693469403992,0.09920634920633332,0.34177215189873417,0.14783260875311086,0.0888888888888,0.791252064624,

MFM1400,0.37931034482750003,0.5641025641025641,0.04867776728783598,0.04545454545455555,0.45569620253164556,0.009594070843313634,0.0,0.664635168305,

Cluster1

Cluster2

Cluster3

Cluster4

**HMS1658**

Cluster0

JHP1654,0.6,0.21428571428571427,0.72737573651609,0.0,0.07407407407407407,0.007689573562431188,0.0,0.723323872126,

WDT1634,0.3333333333335,0.07142857142857142,0.5612630306692854,0.0,0.07407407407407407,0.007988144673132233,0.0,0.635074990653,

DDR1649,0.642857142857,0.6428571428571428,0.42974349935413203,0.0,0.18518518518518517,0.007859448449473677,0.0,0.762951670546,

Cluster1

Cluster2

Cluster3

Cluster4

**HSN0675**

Cluster0

MDS0680,0.45522388059700003,0.5446428571428571,0.12275597279404722,0.12568306010933333,0.4965986394557823,0.07727293132432382,0.054794520548000006,0.768767785149,

Cluster1

Cluster2

Cluster3

MBW1149,0.5,0.008928571428571428,0.007818911850296534,0.0,0.006802721088435374,0.03554643276797502,0.2,0.693982915435,

Cluster4

LMM0167,0.5,0.008928571428571428,0.009251482668434818,0.0,0.006802721088435374,0.00677072533397612,0.0,0.693211509431,

**HXP0976**

Cluster0

Cluster1

Cluster2

Cluster3

KBC1390,0.357142857143,0.5952380952380952,0.5872049784880148,0.0,0.5113636363636364,0.05091383204231564,0.0622222222222,0.821653993437,

Cluster4

NBL1190,0.5,0.6666666666666666,0.5230609030635371,0.0,0.3181818181818182,0.03366934177267629,0.0714285714286,0.868002397571,

**ICB1354**

Cluster0

NAH1366,0.3333333333335,0.2,0.60224,0.0,0.5,0.005267539998332767,0.0,0.677811942,

Cluster1

Cluster2

Cluster3

Cluster4

**ITA0159**

Cluster0

MCP0611,0.3333333333335,0.2,0.544138495637068,0.0,0.2857142857142857,0.19314036707496193,0.2,0.726405559649,

VVG0624,0.4,0.4,0.6314570395958548,0.0,0.42857142857142855,0.09004535049828707,0.1333333333334,0.812330992678,

Cluster1

Cluster2

Cluster3

Cluster4

**JAL0811**

Cluster0

GER0350,0.346153846154,0.1764705882352941,0.2946014283834716,0.17460317460285715,0.27868852459016397,0.01756019144448894,0.014705882353,0.667417449774,

STH0353,0.709677419355,0.43137254901960786,0.10049558338062706,0.05844155844157142,0.14754098360655737,0.004188940274169487,0.0,0.63128503596,

Cluster1

Cluster2

Cluster3

Cluster4

**KSS1005**

Cluster0

Cluster1

GPO1020,0.5,0.01098901098901099,0.23562263984751083,0.3333333333333333,0.00101010101010101,0.0036246414171655947,0.0,0.696270261011,

Cluster2

Cluster3

Cluster4

LVF1626

Cluster0

NWP1609,0.4444444444445,0.1702127659574468,0.5038444427967672,0.0,0.136986301369863,0.039533981104743265,0.03333333333333333,0.713783797746,

ZAD1621,0.5,0.2765957446808511,0.484028494676355,0.0,0.1780821917808219,0.005037689397778164,0.0,0.758362353046,

TAG1610,0.4473684210525,0.36170212765957444,0.4320971392926522,0.0,0.2876712328767123,0.04214487781008366,0.023809523809499997,0.753297863582,

VCF1602,0.3506493506495,0.5744680851063829,0.41583637467042334,0.0,0.684931506849315,0.055486930117612736,0.049999999999999996,0.775833977431,

TNB1616,0.622950819672,0.8085106382978723,0.43648035340105285,0.0,0.3150684931506849,0.04936117199204077,0.036231884058,0.817209940226,

RBC1624,0.51764705882355,0.9361702127659575,0.4105148392051605,0.0,0.5616438356164384,0.07274718052176374,0.06910569105683333,0.894654464381,

Cluster1

Cluster2

Cluster3

Cluster4

CHP1711,0.5,0.02127659574468085,0.24865425965744792,0.0,0.0136986301369863,0.003692252826846532,0.0,0.695814168468,

MCP0611

Cluster0

CKP0630,0.6666666666665,0.5,0.011919467424133197,0.0,0.015384615384615385,0.011152633300291915,0.0,0.616597095261,

Cluster1

Cluster2

Cluster3

Cluster4

MDS0680

Cluster0

Cluster1

Cluster2

Cluster3

Cluster4

MGB1235

Cluster0

Cluster1

Cluster2

Cluster3

Cluster4

NAH1366

Cluster0

Cluster1

Cluster2

Cluster3

Cluster4

OKM1092

Cluster0

HSF1115,0.6,0.1764705882352941,0.005597964696601479,0.0,0.058823529411764705,0.5270976344195779,0.3333333333333333,0.671125922976,

UAM1108,0.2,0.11764705882352941,0.006037450935099551,0.0,0.23529411764705882,0.14482365803710415,0.08333333333333333,0.585123403232,

ZHB1104,0.53846153846155,0.4117647058823529,0.10129886299935315,0.10714285714275,0.1764705882352941,0.20073913755358921,0.12962962963,0.744979393304,

Cluster1

Cluster2

Cluster3

Cluster4

OSS1463

Cluster0

Cluster1

Cluster2

Cluster3

Cluster4

RRS0056

Cluster0

Cluster1

Cluster2

Cluster3

Cluster4

SIS0042

Cluster0

MAF0467,0.9,0.16666666666666666,0.0035609628284293993,0.0,0.009174311926605505,0.006658915816819547,0.0,0.513407067458,

Cluster1

Cluster2

Cluster3

Cluster4

ESP1198,0.875,0.12962962962962962,0.17251363121155103,0.16326530612285714,0.009174311926605505,0.006222803366432107,0.0,0.54862606035,

SNK1280

Cluster0

NAO1281,0.345238095238,0.5272727272727272,0.09125881292217239,0.0935960591132857,0.8870967741935484,0.003375014953901917,0.0,0.671454592259,

Cluster1

Cluster2

Cluster3

Cluster4

IHC0561,0.5,0.01818181818181818,0.0035527746953296717,0.0,0.016129032258064516,0.0017085361078876063,0.0,0.693193255967,

IVS1411,0.5,0.01818181818181818,0.00654057862037148,0.0,0.016129032258064516,0.0034779693662597274,0.0,0.693234684677,

TMT0851

Cluster0

JSB0860,0.5,0.17543859649122806,0.004481739934004949,0.011111111111111112,0.00671591672263264,0.003621891519378966,0.0,0.694526185408,

FKH0864,0.4347826086955,0.17543859649122806,0.046069339236804736,0.044444444444444446,0.008730691739422432,0.04916078583680783,0.042735042735,0.669646029416,

BMR0865,0.3170731707315,0.22807017543859648,0.03863736505593556,0.025641025640999997,0.01880456682337139,0.003099897670416745,0.0,0.613859012644,

DCA0857,0.3507462686565,0.8245614035087718,0.04820773884955058,0.05200945626477777,0.05842847548690396,0.01993056898457924,0.020434227330777776,0.665952854099,

Cluster1

Cluster2

Cluster3

Cluster4

TNB1616

Cluster0

NWP1609,0.3333333333335,0.13333333333333333,0.07825760648454741,0.08333333333333333,0.013029315960912053,0.09543193003278963,0.20833333333333331,0.626999385036,

ZAD1621,0.269230769231,0.23333333333333334,0.14899724374929993,0.3333333333333333,0.030944625407166124,0.06374031857020399,0.08771929824566667,0.647706892414,

TAG1610,0.311111111111,0.4666666666666667,0.13370717817250566,0.21428571428566667,0.050488599348534204,0.052090090382805515,0.04301075268833333,0.69068248549,

VCF1602,0.33928571428550003,0.6333333333333333,0.1299120952764493,0.24561403508766666,0.06026058631921825,0.12020878666354955,0.153153153153,0.744715149166,

Cluster1

KEW0198,0.5,0.03333333333333333,0.006641604941795411,0.0,0.0016286644951140066,0.011916665168054381,0.0,0.693267570853,

Cluster2

Cluster3

Cluster4

TRC1838

Cluster0

RMB1821,0.8,0.09302325581395349,0.5351249551596318,0.0,0.02,0.008533697124748094,0.0,0.582644555903,

FDS1841,0.75,0.20930232558139533,0.45047440635110436,0.0,0.06,0.11623304798983028,0.08333333333325,0.633959707296,

Cluster1

Cluster2

Cluster3

Cluster4

VCF1602

Cluster0

NWP1609,0.5,0.19047619047619047,0.06873107705909869,0.10714285714285714,0.007692307692307693,0.007966280025617244,0.0,0.709788480863,

ZAD1621,0.607142857143,0.40476190476190477,0.04754659898559581,0.050420168067285706,0.010576923076923078,0.0069987470129155285,0.0,0.661720784774,

TAG1610,0.3953488372095,0.40476190476190477,0.06024924045881621,0.05882352941171428,0.025,0.06624240631234313,0.038461538461499994,0.668573897309,

Cluster1

Cluster2

Cluster3

Cluster4

WDT1634

Cluster0

JHP1654,0.5,0.25,0.02174693567685124,0.0,0.11764705882352941,0.007338789500245037,0.0,0.696292289974,

Cluster1

Cluster2

Cluster3

Cluster4

ZIE0741

Cluster0

Cluster1

Cluster2

Cluster3

CDO0684,0.7,0.07216494845360824,0.5142503354094301,0.0,0.02112676056338028,0.24734764426907205,0.4,0.625680622614,

Cluster4

2018年10月26日星期五

下午4时17分

通过实验分析，我们发现如果仅仅考虑ER比值大于0小于1的用户（MinMax后），那么30个跳槽用户中有6个用户是不存在离职用户的：

不存在正常通信的离职邮件联系人。。 CKP0630

不存在正常通信的离职邮件联系人。。 GWG0497

不存在正常通信的离职邮件联系人。。 MDS0680

不存在正常通信的离职邮件联系人。。 MGB1235

不存在正常通信的离职邮件联系人。。 NAH1366

不存在正常通信的离职邮件联系人。。 OSS1463

不存在正常通信的离职邮件联系人。。 RRS0056

CKP0630

Cluster0

Cluster1

Cluster2

Cluster3

Cluster4

RDP1751,0.0,0.0,0.0,0.0,0.14285714285714285,0.0075519706460285595,0.0,0.000654143392704

2018年10月26日星期五

下午10时22分

上述七个用户中只包含ER比要么为1要么为0的单向通讯用户，因此为了坚持出所有的类似用户，我们假设单向通讯也体现了重要的relationlevel。

为了进一步合理的体现ER、单词平均通信量的不同，依旧假设均衡通信最优，而发送要比单纯接收更能体现关系主动性，因此我们修改了公式：



这里的关键是将ER作为一个系数倍数化了关系。

继续进行上述实验

结果实验排序依旧不理想，比如CKP0630只有一个偏远的Cluster4级别的单纯接收用户，可见其relationship并不单纯由邮件决定的，而是决定了邮件通讯；

Insider\_2: 4 : LVF1626 : 1.41076374345

Insider\_2: 249 : HIS1394 : 0.527866195936

Insider\_2: 278 : TNB1616 : 0.489879664328

Insider\_2: 292 : HMS1658 : 0.479828844144

Insider\_2: 306 : ITA0159 : 0.461363895148

Insider\_2: 696 : CIF1430 : 0.238437152595

Insider\_2: 702 : OKM1092 : 0.237548650906

Insider\_2: 759 : TRC1838 : 0.213725833635

Insider\_2: 806 : HXP0976 : 0.20168628672

Insider\_2: 837 : HSN0675 : 0.189844366103

Insider\_2: 857 : CHP1711 : 0.183417862982

Insider\_2: 920 : JAL0811 : 0.162823427839

Insider\_2: 1017 : TMT0851 : 0.13087280797

Insider\_2: 1024 : DCC1119 : 0.129048678309

Insider\_2: 1054 : ICB1354 : 0.121589255352

Insider\_2: 1057 : VCF1602 : 0.121029569017

Insider\_2: 1203 : SNK1280 : 0.0865353179562

Insider\_2: 1296 : BYO1846 : 0.0682071769788

Insider\_2: 1430 : NAH1366 : 0.0448574846602

Insider\_2: 1504 : ZIE0741 : 0.0344187785259

Insider\_2: 1642 : SIS0042 : 0.0161594395164

Insider\_2: 1730 : MCP0611 : 0.008872161012

Insider\_2: 1749 : KSS1005 : 0.00705204975369

Insider\_2: 1759 : WDT1634 : 0.00630000895368

Insider\_2: 1876 : OSS1463 : 0.00145481362209

Insider\_2: 1879 : RRS0056 : 0.00135305279953

Insider\_2: 1909 : GWG0497 : 0.000791912548291

Insider\_2: 1956 : MGB1235 : 0.00018288656574

Insider\_2: 1958 : MDS0680 : 0.000173859288734

Insider\_2: 1960 : CKP0630 : 0.000130828678541

如果单独考虑规范邮件，筛选效果也不理想：

Insider\_2: 41 : LVF1626 : 2.46085272673

Insider\_2: 45 : HIS1394 : 2.37112648428

Insider\_2: 188 : TNB1616 : 1.51159738979

Insider\_2: 477 : TMT0851 : 0.856461650789

Insider\_2: 586 : SNK1280 : 0.719710511701

Insider\_2: 617 : HMS1658 : 0.676104118385

Insider\_2: 650 : JAL0811 : 0.643099166017

Insider\_2: 679 : VCF1602 : 0.603032599255

Insider\_2: 691 : CHP1711 : 0.59021376565

Insider\_2: 726 : HSN0675 : 0.557141466179

Insider\_2: 778 : ITA0159 : 0.502062848809

Insider\_2: 834 : OKM1092 : 0.438825855292

Insider\_2: 835 : CIF1430 : 0.438359909872

Insider\_2: 908 : BYO1846 : 0.375536224562

Insider\_2: 947 : TRC1838 : 0.341828549917

Insider\_2: 1174 : HXP0976 : 0.166576092066

Insider\_2: 1177 : DCC1119 : 0.165037339837

Insider\_2: 1281 : ICB1354 : 0.113789573511

Insider\_2: 1425 : SIS0042 : 0.0491005799336

Insider\_2: 1480 : ZIE0741 : 0.033392049361

Insider\_2: 1558 : MCP0611 : 0.00875420707227

Insider\_2: 1678 : OSS1463 : 1.02137494457e-05

Insider\_2: 1769 : CKP0630 : 0.0

Insider\_2: 1824 : GWG0497 : 0.0

Insider\_2: 1881 : KSS1005 : 0.0

Insider\_2: 1902 : MDS0680 : 0.0

Insider\_2: 1904 : MGB1235 : 0.0

Insider\_2: 1917 : NAH1366 : 0.0

Insider\_2: 1957 : RRS0056 : 0.0

Insider\_2: 1984 : WDT1634 : 0.0

因此，假设不成立，不能由邮件通讯反推到relationship