# 技术模式

1. 行为动作分析
2. 言论分析
3. 网站搭建
4. Netlogo实证模拟
5. 综上投入运用

## 一、概览

本部分主要介绍本项目组中新技术的原理简要以及运用方式。通过可穿戴设备数据分析、文本分析、基于NetLogo技术的实证模拟、网站搭建等方式，，降低本项目的成本、提升资源利用效率，为本项目的成功开展保驾护航。

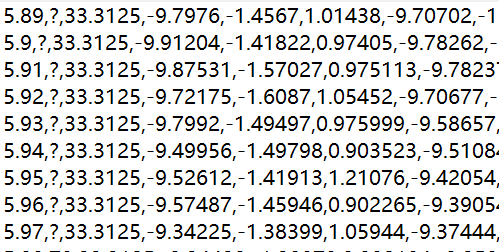
## 二、技术运用

### 1.基于监督学习的行为动作分析

可穿戴设备可以使我们有效的利用传感器获取人类的生理数据，本项目的目标是根据传感器收集到的数据对人体的坐姿进行基本分析。

现在我们收集了A，B，C，D，E一共五位用户的可穿戴设备上的传感器数据，每位用户的数据集包含一个特征文件和一个标签文件。

特征文件中每一行对应一个时刻的所有传感器数值，标签文件中每行记录了 和特征文件中对应时刻的标记过的用户姿态，两个文件的行数相同，相同行之间互相对应。



特征文件数据示例

特征文件一共包含41列数据：第一列为时间戳，第二列心率。后续30多行的数据是由三个传感器依次记录的数据：

在传感器1对应的13列数据特征中，包含：1项温度数据、3项一型三轴加速 度数据、3项二型三轴加速度数据、3项三轴陀螺仪数据和3项三轴磁场数据。

人体的温度数据可以反映当前活动的剧烈程度，一般在静止状态时，体温趋 于稳定在36.5度上下；当温度高于37度时，可能是进行短时间的剧烈运动，比如跑步和骑行。

在数据中有两个型号的加速度传感器，可以通过互相印证的方式，保证数据 的完整性和准确性。通过加速度传感器对应的三个数值，可以知道空间中x、y、 z三个轴上对应的加速度，而空间上的加速度和用户的姿态有密切的关系，比如用户向上起跳时，z轴上的加速度会激增。

陀螺仪是角运动检测的常用仪器，可以判断出用户佩戴传感器时的身体角度是水平、倾斜还是垂直。直观地，通过这些数值都是推断姿态的重要指标。

磁场传感器可以检测用户周围的磁场强度和数值大小，这些数据可以帮助我们理解用户所处的环境。比如在一个办公场所，用户座位附近的磁场是大体上固定的，当磁场发生改变时，我们可以推断用户的位置和场景发生了变化。

标签文件包含如下：每一行代表与特征文件中对应行的用户姿态类别。 总共有0-24共25种身体姿态，如，无活动状态，坐态、跑态等。标签文件作为 训练集的标准参考准则，可以进行特征的监督学习。我们可以选定某种分类模型（或者说是 算法），通过使用训练数据进行模型学习，然后对每个测试样本给出对应的分类结果。

## 算法流程

1. 读取文件，对数据进行预处理。
2. 创建分类器，对数据开展训练。
3. 计算模型准确率和召回率，评估模型。

具体实现：

1. 调用pandas的read\_table（）函数对文件进行读取。
2. 将前四个用户数据作为训练集，最后一个用户E作为测试集。故使用train\_test\_split打乱数据，将test\_size赋值为1（因为设置为0，会出现异常），即打乱训练集数据即可。
3. 将训练集送入K近邻、决策树、贝叶斯模型进行训练，并进行预测，得出其准确率、召回率、F1值。

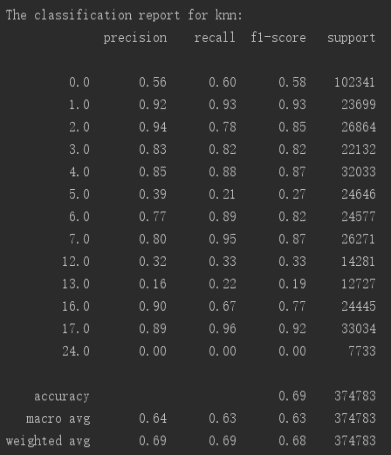
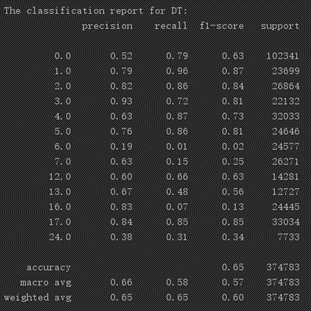
注：

**1.精确率(Precision）**是指在所有系统判定的“真”的样本中，确实是真的的占比，就是TP/(TP+FP)。

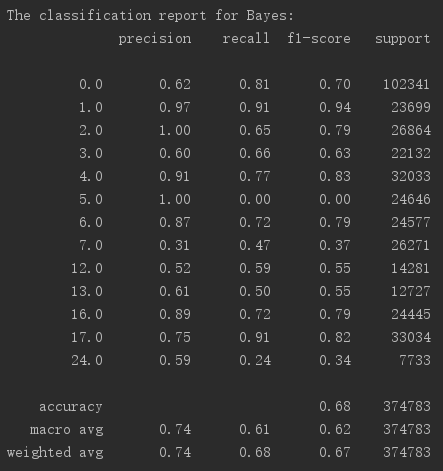
**2.召回率（Recall）**是指在所有确实为真的样本中，被判为的“真”的占比，就是TP/(TP+FN)。

3. **F1值**是为了综合考虑精确率和召回率而设计的一个指标，一般公式为取P和R的harmonic mean:2\*Precision\*Recall/(Precision + Recall)。

结果如下：

KNN DecisionTree



GaussianNB

结果对比：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 准确率 | 召回率 | F1值 |
| KNN | 0.69 | 0.69 | 0.68 |
| 决策树 | 0.65 | 0.65 | 0.60 |
| 贝叶斯 | 0.74 | 0.68 | 0.67 |

结论如下：

1.从准确度的角度衡量，贝叶斯分类器的效果最好

2.从召回率和F1值的角度衡量，k近邻效果最好

3.贝叶斯分类器和k近邻的效果好于决策树

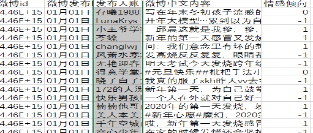
通过上文的分析，我们可以利用上文的训练结果，对其投入应用，通过时时刻刻分析老人身上的可穿戴设备传来的数据，我们可以实时监测老人身体姿态，监测其身体状态，保证其生命安全，

在项目发展到到一定规模后，我们还将把可穿戴设备发给购买产品但是并未年老的客户，通过监测其坐姿研究其生活习惯。通过研究客户的生活习惯，我们会适当的为拥有健康生活习惯（比如驼背坐姿较少的顾客）在保费上提供适当的优惠，帮助并督促客户形成健康的生活习惯，对提升客户健康观念，提升客户身体素质水平有很大助力。同时，通过几年累月的生活习惯观测，我们还可以通过分析，得出客户的慢性病或者可能罹患的疾病，对顾客有针对性的订制护理方案以及保养方案，使每一个顾客都能享受到对身体独一无二的定制化的照顾。

### 2.文本分析

在用户给我们的提交的反馈中，为了避免不实的反馈情况，我们选择了对其反馈意见进行文本分析，鉴于当前技术原因，在此只对情感分析模块做一定演示。此处研究的短文本情感分类是根据文本内容所体现的用户意见的情感极性, 将带有相同特定情感倾向的短文本归为一类, 即文本情感分类。目前短文本情感分类主要采取统计自然语言处理、情感语义特性两种方式。此处所采取的方式是利用jieba分词库对我们已经获取的文本进行分词，在调用函数进行向量化，向量化之后利用已知的lable数据对当前数据做一个监督学习，训练模型之后再对预测集进行预测，此处衡量预测结果的不再使用准确率，而是采用Macro-F1值，对文本分析结果予以分析。

本文训练和预测所采用的数据是来自于北京市经济和信息化局所举办的情感分析分析的比赛数据：数据集依据与“新冠肺炎”相关的230个主题关键词进行数据采集，抓取了2020年1月1日-2020年2月20日期间共计100万条微博数据，并对其中10万条数据进行人工标注，标注分为三类，分别为：1(积极)，0(中性)和-1(消极))。故而在此采用的数据仍然是带有标签的，所以仍然采用较为初步的监督学习对其进行归类。



首先，利用jieba’分词对已经获取到的数据进行分词，接着调用哈工大的停用词表，再通过设置参数过滤调过于平常的词、停用词和过于特殊的词。

情感分类器主要有支持向量机、K最近邻和朴素贝叶斯。支持向量机对处理样本量少的数据效果较好, 但大样本时优势并不明显, 且这种非线性算法的计算复杂度较高, 也不适合在大样本数据上做训练。贝叶斯分类是根据某对象的先验概率, 利用贝叶斯公式得出后验概率, 以最大后验概率的类作为该对象所属的类。而KNN分类器是根据距离度量个体间的差异性, 将距离相近的归为一类。本项目决定在此采用的朴素贝叶斯分类器。

在此基础上将数据投入其中进行学习，获取结果后再对预测集进行预测，预测完后验证其macroF1值，在此本文的MacroF1值为0.66左右。

在此基础上，我们会利用相似技术对客户的反馈进行分析，准确把握客户在意见反馈之下的情感变化，及时修正我们的服务方向和服务配置，增强我们的纠错能力和点对点能力。

### 三、网站搭建

本项目组依然会建立自己的官网，利用网络来进行宣传本项目，同时招募、联系、评分、反馈等系统也都将于本网站上进行实现。

同时我们会建立相关数据库，利用好数据库技术，妥善存储志愿者和老人的信息。并对其加以管理，与时间银行、服务订制等模块相结合，利用好数据库技术。

网站具体实现细节和功能可以参见本策划书——网站架构环节。

### 四、基于netlogo多主体技术的实证分析

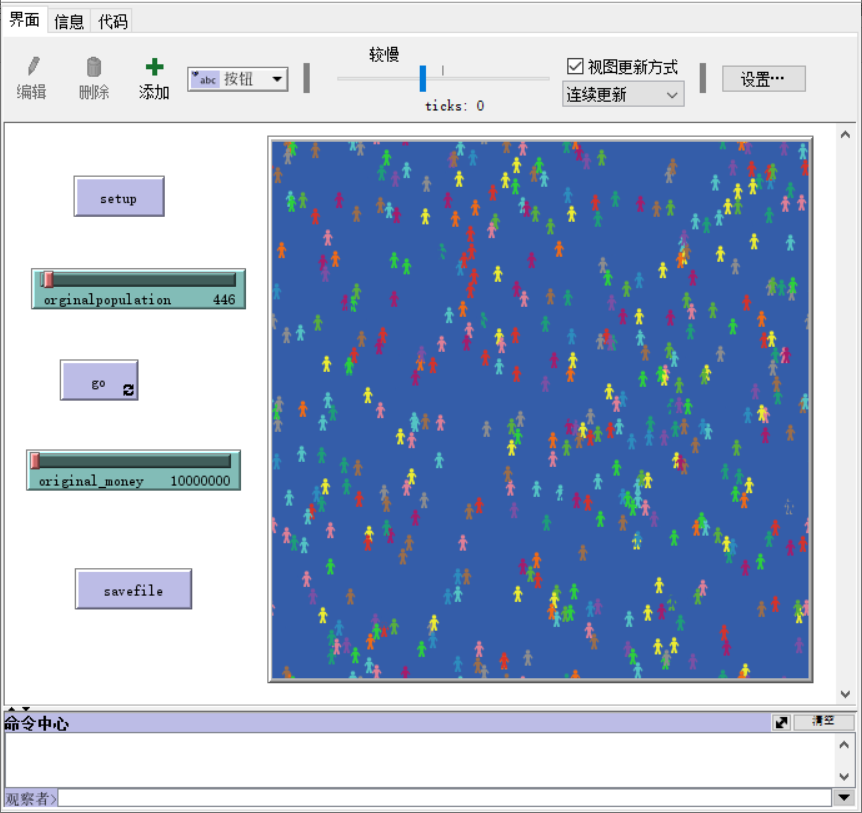
NetLogo是继承Logo语言的一款编程开发平台,但它又改进了Logo语言只能控制单一个体的不足,它可以在建模中控制成千上万的个体,因此,NetLogo建模能很好地模拟微观个体的行为和宏观模式的涌现及其两者之间的联系。NetLogo是用于模拟自然和社会现象的编程语言和建模平台,特别适合于模拟随时间发展的复杂系统。

为了更好的验证本项目的可行性，所以使用net logo多主体模拟软件对人类社会的自然生老病死进行模拟，并且给出各年龄段的死亡人数。setup初始化当前界面并定义turtle，通过滑块调整turtle数量。

首先，我们定义好turtle遵循的规则：

1. 设置turtle每走一步就增加一岁，初始的turtle将会都从二十岁开始。
2. 每个turtle都有性别，假设初始一半为男，一半为女。其中，男为1，女为-1.
3. 每一岁都有固定的死亡率，死亡率参考《中国人身保险业经验生命表》。
4. 设定起始turtle的起始财富相同，开始的时候所有人的财富是相同的，若二人相遇，会随机分配货币（即充当花销部分）。且财富量每年会随着gdp的增长速度而增加。
5. 身为经济人，每个经纪人会有一定的存款率s1，即他在交易时保留m1s1的金额作为存款，以剩下的来作为分配。
6. 设定每个经纪人有生子的概率（生育率参考国家统计局统计年鉴）

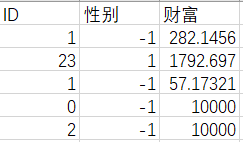
由此对人类社会进行模拟，并且导出数据。



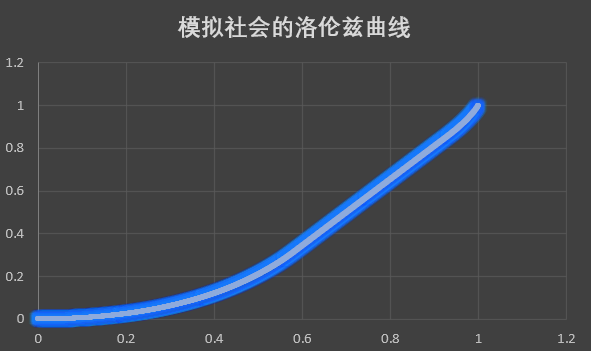
图：对人类社会的模拟

如图，在本次模拟中，orginalpopulation滑块可以设定社会的起始人口数量，originnal\_money滑块则可以设定社会的财富总量。考虑到GDP增长和人口增长的关系，所以在这里我们将社会总财富的增长与turtle的总数相挂钩，完善我们的社会模型。

使用save\_file函数导出数据，导出的数据如下：



此时，该社会的人口数已经达到了5000多，为了考量对人类社会模拟的拟真程度，让我们考量一下这个小社会的部分特征：



如图是该社会的洛伦兹曲线，估算其基尼系数，大概处于0.4左右。此时社会才仅仅到达5000余人的规模（碍于笔者计算机性能，后期社会模拟程度所需turtle更多，暂时难以模拟）。

一般居家养老的老人，考虑到其与外界的商业往来并不存在中介、大幅度优惠等情况，故可以在此粗略估计为居家养老者与年轻人的花销方式并无不同。此时的人均财富仅为5664.29/元左右，如果才用一般养老保险制度，设定其年轻时（20-60之间）每年支出10000元于养老保险之上，待其老区之后，每年可获得20000元/人，此时的社会人均总财富为6438.049元/左右。如果采用本项目组的服务，因为本项目组的服务特性，价格较低，同时志愿服务的存在，导致部分人并不会购买产品，而是通过大力做志愿等行为换取养老服务或者缩短购买产品时间，故而设定其为扣除5000/年/人，待其老后设定为每人每年获取10000元左右的价值回报，检测社会人均总财富，此时人均总财富达到6864.052元，故而可得出我项目对社会的正向作用。同时，提高人均总财富，也有利于潜在客户市场承接我们的产品。

### 五、综合

综上所述，本项目在运营得过程中，我们会灵活使用可穿戴设备，利用机器学习技术对其进行数据分析，灵活掌握客户的身体姿态，确保其生命安全和身体健康，督促其健康生活习惯的养成，使每个客户享受到独有得订制服务。

同时利用文本分析，妥善的保证我们志愿服务的质量，及时收集、提取意见，改进志愿服务，为大家带来更好的志愿服务体验，为社会带来更好的正向性。

利用网站搭建好平台，使用netlogo仔细评估每一个决策，确保本项目组的运营日趋完美。