

中山大学移动信息工程学院本科生实验报告

(2015 年秋季学期)

课程名称: Artificial Intelligence

任课教师: 饶洋辉

年级	13 级	专业 (方向)	移动信息工程
学号	13354485	姓名	朱琳
电话	13726231932	Email	280273861@qq.com

一 代码实现部分——1NN

1. 数据结构

```
struct Files { //每个文本的数据
    float vector[COL];
    int shares; //每个文本被分享的次数
};

float dis[TEST][TRAIN];
```

2. 文件的读取

```
void readTrain() {
    ifstream fin(trainData);
    float v;
    char ch;
    string w;
    getline(fin, w);
    int col = 0;
    while (fin >> v) {
        if (col < num_of_col) {
            trains[num_of_trains].vector[col++] = v;
            fin >> ch;
        } else {
            trains[num_of_trains].shares = v;
            col = 0;
            num_of_trains++;
        }
    }
    cout << "num_of_trains=" << num_of_trains << endl;
    fin.close();
}
```

3. 计算距离, 这里采用余弦距离

```
void get_dis() {
    for (int ttRow = 0; ttRow < num_of_tests; ttRow++) { //每个测试文本
        for (int tnRow = 0; tnRow < num_of_trains; tnRow++) { //每个训练文本
            dis[ttRow][tnRow] = 0;
            float distance = 0;
            for (int k = 0; k < num_of_col; k++) {
                distance += abs(
                    trains[tnRow].vector[k]*tests[ttRow].vector[k]);
            }
            dis[ttRow][tnRow] = distance;
        }
    }
}
```

4.1NN

因为是余弦距离，所以距离最大的相似度最大。

```
void oneNN() {  
    for (int ttRow = 0; ttRow < num_of_tests; ttRow++) { //对于每个测试文本  
        //float min = INT_MAX;  
        float max=INT_MIN;  
        for (int tnRow = 0; tnRow < num_of_trains; tnRow++) {  
            if (dis[ttRow][tnRow]>max) {  
                max = dis[ttRow][tnRow];  
                tests[ttRow].shares=trains[tnRow].shares;  
            }  
        }  
    }  
}
```

二 实验结果

这两天比较忙，还没来得及实现其他的方法，1NN 的效果不太好，只能达到 0.56，开始的时候被文件读写坑了很久，原来当 txt 打开还有一个逗号要读。然后下周主打决策树，NB 也会实现一些，现在我的进度有些慢了，前几天被一堆事情拖着。