“拍照赚钱”APP 中任务定价问题分析及求解

许祖铭

（兰州理工大学计算机与通信学院 15 级物联网工程，甘肃　兰州　730050）

摘　要：“拍照赚钱”的任务定价问题选自 2017 年高教社杯全国大学生数学建模比赛 B 题，其中“拍照赚钱”是一款自助式服务 APP 提供的主要业务。任务定价问题本质是一个优化问题，如何在付出最低成本的同时任务的完成率最高是本文的研究重点。

关键词：任务定价；优化问题；最低成本；最高完成率

中图分类号：TP139

根据题目信息，可知“拍照赚钱”APP 中的任务定价、任务分配不尽合理。部分任务无人问津，某些地区的任务完成率异常的低；而有些地区的任务完成率则接近 100%，任务的分配与定价不合理成为影响 APP 发展和推广的主要问题。通过对 APP 统计数据进行处理和分析，观察任务的定价规律，结合各种因素分析任务未完成的原因，建立数学模型对原有的定价机制进行改进。并结合会员的分布情况，信誉值等因素，给出优化的任务分配方案，实现以最低成本取得更高任务完成率。

1　问题提出

1）研究附件一中数据，分析任务定价规律及未完成的原因。

2）为附件一的任务设计新的定价方案，并和原方案进行对比试验。

3）实际情况下，多个任务可能因为位置比较集中，导致用户会争相选择，一种考虑是将这些任务联合在一起打包发布。在这种考虑下，如何修改前面的定价模型，对最终的任务完成情况又有什么影响？

4）对附件三中的新项目给出你的任务定价方案，并评价该方案的实施效果。

2　问题分析及求解

2.1　问题一分析及求解

附件中有 835 项任务记录，包括任务地点的经度、纬度、任务标价和完成情况。首先，需要明确，任务的位置是如何影响其定价的。可以利用 EXCEL 以及 BDP 应用程序将全部任务的信息以散点形式在地图上表现出来，利用不同颜色区分完成与未完成的任务，利用同一颜色的深浅程度来表示价格高低，以此研究任务位置对定价和完成率的影响。实验产生的散点图呈现出若干个聚集区，而价格则大致随（与聚集中心的）距离的递增而衰减。

为验证所得结果，使用 SPSS 的 K-means 聚类分析[1]验证。首先，将任务划分为四个区域，并设置每个区域的中心点。然后，对任务价格与任务到聚类中心点[2]的距离之间的关系进行研究，采用多项式回归进行拟合分析[3]。最后，将未完成的任务单独提　取，从任务的分布和定价、当地的会员信息、任务分布区域的经济状况、交通因素等方面，分析任务的完成情况。

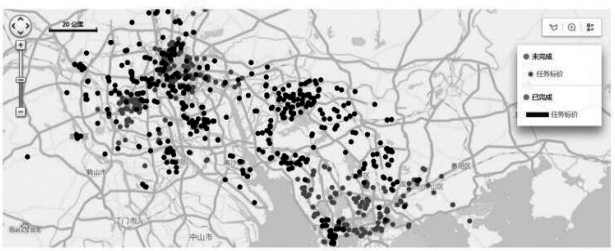

图 1　未完成和已完成任务分布图

图 1 中展示了任务依据坐标位置在地理上的呈现，其中红色圆点表示未完成任务，黑色表示已完成任务，通偶对该图的观察，发现黑色的点即完成的任务呈聚集状态，并且聚集中心主要分布在城市中心。说明距离市中心的距离直接影响到任务的完成率，并且呈负相关。

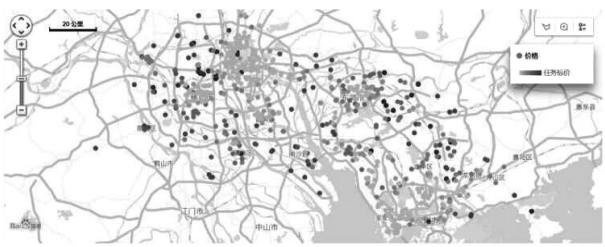

图 2　价格梯度分布图

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

图 2 中展示了任务依据坐标位置在地理上的呈现，通过红色的深浅表示任务定价的变化，红色越深表示任务的定价越高；通偶对该图的观察，发现任务的定价呈聚集状态，并且聚集中心主要分布在城市中心，随着距离市中心的距离不断增加，任务的定价逐渐降低，红色颜色变浅。说明距离市中心的距离直接影响到任务的定价，并且呈现负相关。

由于东莞市任务完成率为 100%，说明任务定价过高，深圳市任务完成率仅为 22%，反映了定价偏低；而另外两个城市的完成率则较为适中。所以选取广州市与佛山市的数据来分析（任务点到聚类中心[1,2]的）距离以及与任务价格的关系。使用 MATLAB 拟合线性回归函数[3]P 与距离 d 的函数关系如下，图 3 为拟合函数曲线。

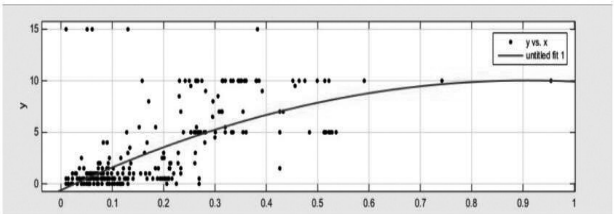

图 3　拟合函数

拟合函数：P (d)=−12.86d2+23.40d−0.62，R12=0.435

任务未完成原因分析：

1)任务地理位置的影响；同样价格的任务，靠近市中心，则完成的机率较大；位于市郊，则完成的机率很小。

2)任务价格的影响；因为人们完成任务是以获得酬劳为目的的。

3)会员地理位置分布的影响；比如，在佛山市的一部分区域，任务多的地方会员数少，会导致部分任务没人完成。

4)会员信誉值的影响；在部分任务量偏多的地区，信誉值偏低的会员数多，抢占优质会员资源。

5)当地经济发展水平的影响；

6)交通因素的影响；观察到靠近市中心也存在大量未完成的任务，排除价格因素，原因可能是没有交通直达，需要步行，即使价格合理，也没有会员去完成。

2.2　问题二分析及求解

基于问题一分析，可知任务位置坐标、任务定价、会员坐标、会员信誉值、会员配额是任务完成的主要影响因素。该定价模型本质是非线性回归模型[4]。我们从任务定价出发，数据集记录可以划分为任务定价合理和不合理两类，把定价转化为二分类问题[5]，以此研究多因素对任务完成率的关系，找出价格定价是否合理。迭代价格范围，找出在模型下可以完成的最低价格作为定价。支持向量机(SVM)[6]是在分类与回归分析中分析数据的监督式学习模型[7]。SVM 模型的基本流程图如图 4 所示。

图 4　SVM 模型整体流程

给定一组训练实例，每个训练实例被标记为属于两个类，即价格定价合理和不合理，SVM 训练算法创建一个将新的实例分配给两个类别之一的模型，使其成为二元分类器。

通过使用核函数，SVM 可以有效地进行非线性分类，将其输入隐式映射到高维特征空间中。高维空间中的线性决策界实际上是低维空间的非线性决策界，这个非线性决策界是线性分类器找不到的，而核函数可以有效解决该问题：

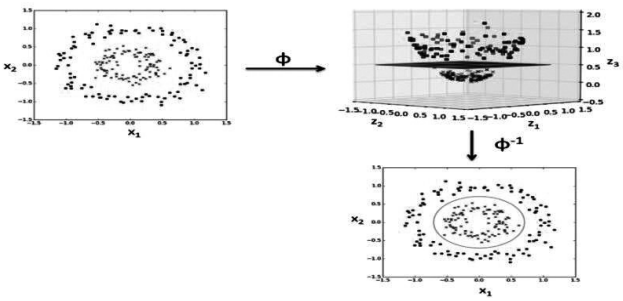

图 5　SVM 的核方法

图 5 显示的是采用 SVM 方法对数据在三维空间的展示，其中蓝色的点显示的是定价合理的点，红色部分表示的是不合理的点，采用核函数可以有效的划分定价的两个区域。

2.2.1　数据处理

综合考虑任务表和会员信息表，构建特征向量，作为建模的数据源，并对会员表数据进行处理。以会员信誉值为权重，重新调整限额数，我们在表中新加了一列即会员有效限额数。第 n 个会员的有效限额数为：

$$T_n = \begin{cases} \sum_{k=1}^{N} x_k \cdot \frac{y_n}{\sum_{i=1}^{N} y_i} & (T_n < x_n) \\ y_n & (\text{其他情况}) \end{cases} \tag{1}$$

其中，n 表示第 n 个会员，N 表示任务数；T_n 为

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

会员的有效限额数，k 表示任务数目，$k\varepsilon[1,N]$，x_n 表示第 n 个用户的信誉值。y_n 表示第 n 个用户实际的完成任务数；

2.2.2 合并任务表和会员表

在数据合并过程中，以每个任务坐标为中心，对周围1km范围内的会员进行计数并累加会员的有效限额数。坐标保留四位小数，新数据保存为五列：经度、纬度、任务价格、累计会员有效限额、是否完成。

2.2.3 利用SVM模型求解

常用的一个核函数是Radial Basis Function kernel(RBF核)，也称为高斯核。

模型训练参数设置为：(1) 训练集/测试集比例：1:1；(2)参数gamma范围0-50的整数；(3)参数 C范围：1-50整数。最终得出的最佳参数为：gamma=6，C=1，模型评价参数：precision=0.66，recall=0.66，f1-score=0.62，support=418。

2.2.4 与原方案进行比较

根据模型二得到新的任务定价方案。原方案完成522个任务，任务完成率0.625，完成任务花费总金额为36446元，任务价格平均值为70元；现方案预计完成785个，任务完成率0.94，预计花费金额为50088，即任务平均价格为63元。该方案与原方案比较，任务完成率提高30个百分点，单个金额降低7元均优于原方案，对比结果见表1。

表1 原方案和改进方案结果对比

名字	总任务数	完成方案数	完成率	任务总花费	任务平均花费
原方案	835	522	0.625	36446	70
改进方案	835	785	0.94	50088	63

模型二为非线性回归模型[4]，不仅考虑了任务经纬度因素，也考虑会员分布、会员信誉度以及预定任务限额等因素。

2.3 问题三分析及求解

考虑到实际情况，需要将任务进行打包。首先，通过EXCEL按信誉值对会员进行排序，并通过有效限额数公式，将信誉值转换为会员可接收的任务数。以完成任务为中心，通过任务分布密度和预定任务限额两个参数，将范围内的未完成 任务打包。遍历数据集，找出距离完成任务位置半径0.03km内的未完成的任务，打包任务设置每个包可以发布的最大值。由此得到打包任务的总价格以及每个包所含的任务数。考虑以上因素，完善问题二中的SVM模型，使其预测精确度更高，提高任务完成率。

2.3.1 拟合打包任务

打包任务以已完成任务为中心，遍历数据集，找出坐标半径为0.03km内的未完成的任务，打包任务设置每个包可以发布的最大值。由此得到打包任务的总价格以及每个包所含的任务数。部分打包数据见表2。

表2 打包任务的总价格以及每个包所含的任务数

经度	纬度	总价	会员限额数	每个包任务数	完成度
22.576	113.957	328.5	177	8	1
22.549	113.972	335	25	5	1
22.543	113.924	262.5	42	4	1
22.547	113.977	65.5	1	1	1
22.751	114.026	140	2	2	1
22.738	113.818	66.5	8	1	0

2.3.2 SVM支持向量机参数训练

使用SVM[5,6]建模，经训练得到最佳参数C=1;gamma=13，测试结果如下precision=79%，recall=80%，f1-score=75%，support=776，通过打包，完善了问题二中的SVM模型，使其预测精确度更高，提高了任务完成率。

2.4 问题四分析及求解

该问题要求对附件三中的新项目给出任务定价方案，在使用模型的过程中，由于各个维度之间的特征没有统一量纲，而在建立模型的过程中又没有对特征进行归一化处理，因此模型的泛化能力大为降低。在第二个问题建模过程中，抓住了会员信誉和任务量以及时间的线性关系，成功对数据做了处理；但在问题三处理打包任务时，引入的处理条件过于强硬，以至于结果拟合程度剧增。所以可以使用问题二中的模型对新项目给出定价方案。

步骤一：首先打包任务；

（下转第17页）

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

在 PATHLOSS 软件中，当天线设置好了以后，看微波的反射是否影响接收，进入主菜单 Module> Multipath。可以看到图 4 画面。

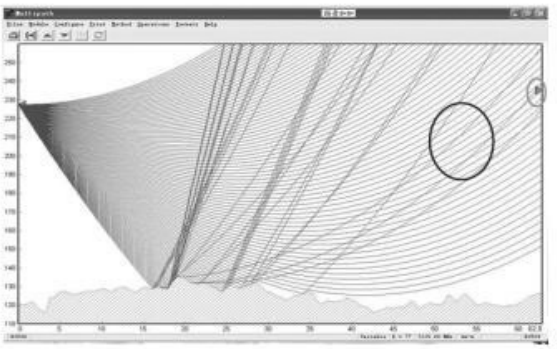

图 4　多径反射有干扰图

当出现图 4 所示的画面时，表示反射对接收有影响。这样，就必须修改天线的高度，直到反射对接收没有影响为止。

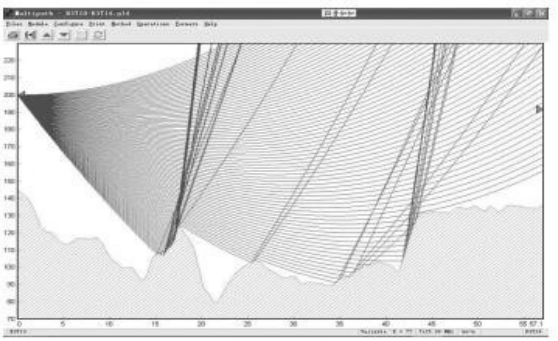

图 5　多径反射无干扰图

图 5 中红色的为反射微波，三角型为微波天线，接受点处避开反射，表示反射对接收没有影响。此时的天线挂高既满足了余隙要求又避开多径反射，是最优天线挂高。

4　结束语

以上为使用 PATHLOSS 软件设计天线挂高，还用此软件实现微波通信设计中的网络预组网规划、容量规划、链路性能计算、频率规划、干扰分析，规划过程中可以依照仿真结果不断做网络调整，并在最后达到微波通信设计指标。线路设计完成，PATHLOSS 软件还可输出打印设计结果图纸和表格，包括网络组网图、站点链路信息、链路径剖面图、链路性能计算报表、频率干扰与高低站冲突报告，为全省微波电路升级改造提供强有力的分析、判断与参考。

参考文献：

[1] 王琦.利用 GOOGLE EARTH 地图进行微波网络规划[J].无线通信技术,2011,3:59-61.

[2] 张旭.微波路由设计和天线挂高计算的探讨[J].河南科技,2007,3:35-36.

[3] 赵黎明.用微机计算微波通信线路中的天线挂高[J].浙江电力,1998,2:64-67.

[4] 丁荣林,李媛.微波技术与天线(第二版)[M].北京:机械工业出版社,2013.

(上接第 20 页)

步骤二:迭代求出任务价格,经测试,新任务共完成 1690 个,花费 108539;平均单个花费 64 元,任务完成率 84.5%。相比单个定价模型效率有所降低。

3　总结

模型优点:利用 SVM 建立数学模型,进行非线性回归分析,综合考虑了经纬度,会员分布,信誉值,任务限额,完成率等多种因素,并充分考虑实际情况。降低了价格,提高了任务完成率,优化了原有定价方案。模型缺点:数据集特征缺乏统一量纲,建立模型的过程中对特征缺少归一化处理,导致模型的泛化能力降低。实验数据中,模型的准确率达到 94%,但对新数据的处理能力有所降低。在第二个问题建模过程中,我们主要研究会员信誉和任务量以及时间的线性关系;在处理打包任务时,引入的处理条件过于强硬,以至于结果过拟合程度剧增。

参考文献：

[1] 张建萍,刘希玉.基于聚类分析的 K-means 算法研究及应用[J].计算机应用研究,2007,(05):166-168.

[2] 薛印玺,许鸿文,李羚.基于样本密度的全局优化 K 均值聚类算法[J/OL].计算机工程与用,:1-6(2017-10-10). http://kns.cnki.net/kcms/detail/11.2127.TP. 20171010. 1536.010.html.

[3] 郑绪枝,杨柱元.基于小波的多项式回归及预测[J].云南民族大学学报(自然科学版),2012,21(03):182-184.

[4] 龚江,石培春,李春燕.巧用 Excel 解决多元非线性回归分析[J].农业网络信息,2011,(01):46-48.

[5] 李欢,王士同.支持向量机的多观测样本二分类算法[J].智能系统学报,2014,9(04):392-400.

[6] 丁世飞,张健,张谢锴,安悦瑄.多分类孪生支持向量机研究进展[J/OL].软件学报,:1-21(2017-07-12).http://kns.cnki.net/kcms/detail/11.2560.TP.20170712.1616.017.html. DOI:10.13328/j.cnki.jos.005319.

[7] 杨南海,黄明明,赫然,王秀坤.基于最大相关熵准则的鲁棒半监督学习算法[J].软件学报,2012,23(02):279-288.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net