

所属类别	2021 年“华数杯”全国大学生数学建模竞赛	参赛编号
研究生组		CM210527

电动汽车目标客户销售策略研究 ——基于附加动量的 BP 神经网络模型分析

摘要

在新能源浪潮下，新能源汽车产业作为战略性新兴产业，其长期的发展趋势确定，但目前存在国内新能源汽车渗透率不足等问题，仍具有广阔的成长空间。消费者在新能源汽车的续航以及实用性领域等方面存在疑虑，因此对新能源汽车销售方面的研究具有很大的价值。本文根据某汽车公司的调研数据，将目标客户对三种品牌车型的满意度得分以及目标客户的个人特征作为购买意愿的影响因素。紧扣购买意愿的驱动因素分析，利用主成分分析方法（PCA）对影响因素进行降维，最终提取了九个主成分，其中电动汽车本身的因素对客户购买意愿的影响最大，其次是客户的收入、驾龄、工作年限等个体特征。本文利用机器学习方法，在 Keras 的深度学习框架下使用多层感知器神经网络模型实现对客户购买意愿的二分类预测。此外为提高 BP 神经网络对目标客户购买行为识别的准确率，本文利用附加动量法建立基于附加动量的 BP 神经网络模型，并依据部分用户的特征预测其购买意愿。综合以上分析，本文提出从学历以及职业角度拓宽市场以进行精准营销，以及从客户满意度出发提高新能源汽车优势的销售策略。

关键词：附加动量的 BP 神经网络；多层感知器神经网络；购买意愿分析

目录

一、引言.....	1
二、描述性分析.....	1
（一）数据预处理.....	1
（二）用户特征描述性分析.....	1
1、目标客户收入差距分析.....	1
2、目标客户购买意愿分析.....	3
（三）目标客户满意度比较分析.....	5
三、主成分分析（PCA）.....	6
（一）主成分分析适用性检验.....	6
（二）求解主成分.....	7
（三）计算主成分得分.....	9
四、神经网络模型.....	10
（一）多层感知器神经网络模型（MLP）.....	10
1、多层感知器神经网络模型（MLP）.....	10
2、神经网络的构建.....	10
3、神经网络结果分析.....	12
（二）BP 神经网络模型.....	12
1、BP 神经网络模型.....	12
2、附加动量的 BP 神经网络的构建.....	14
3、BP 神经网络结果分析.....	16
五、目标客户销售策略.....	16
六、结论与建议.....	17
（一）开启蓝海市场、针对特定用户精准营销.....	17
（二）提高电动汽车产品质量、性能和服务.....	18
（三）对不同品牌目标客户实施精准营销。.....	18
参考文献.....	19
附录一 数据清洗与描述性分析代码.....	20
1、数据清洗 SQL 语句.....	20
2、图 1 代码.....	20
3、图 2 代码.....	21
4、雷达图 Python 代码.....	21
附录二 各因子得分计算过程.....	22
附录三 神经网络代码.....	23
1、Python 多层感知器神经网络模型（MLP）代码.....	23
2、MATLAB 附加动量的 BP 神经网络代码.....	26

一、引言

新能源汽车是指采用非常规的车用燃料作为动力来源（或使用常规的车用燃料、采用新型车载动力装置），综合车辆的动力控制和驱动方面的先进技术，形成的技术原理先进、具有新技术、新结构的汽车。近年来，随着科学技术的更新换代，在新能源浪潮下，新能源汽车产业作为战略性新兴产业，长期成长趋势确定。据《新能源汽车产业发展规划（2021-2035 年）》，到 2025 年，新能源汽车销售量达到汽车新车销售总量的 20%左右。而目前国内新能源车渗透率不足 8%，仍具广阔的成长空间。新能源汽车作为新兴产业，他与传统汽车相比，消费者在一些领域，比如电池能耗等方面对它的接受还需要一定的时间。

神经网络对于数据的判断应用的越来越广泛，本文通过运用人工神经网络以及附加动量的 BP 神经网络算法，对训练集的数据进行测算，提取目标客户中对三种品牌的汽车的满意度以及用户自身特征，最后对测试集的客户进行购买意愿的判断。同时用主成分分析法对影响客户购买的各个因素进行有效信息的提取，从不同的角度对用户行为进行探究，以此为新能源汽车的更新以及销售提供可行性建议。

二、描述性分析

（一）数据预处理

目标客户体验数据包括目标客户对三款品牌电动汽车体验的满意度评价信息、目标客户个人特征信息，共 28 个变量。

本文首先对数据进行了预处理，针对整体满意度得分超过 100 分以及支出占总收入的比例超出 100%的异常值进行了剔除。根据目标客户个人特征调查表中的出生年份信息，得到孩子个数变量中存在缺失值的个体为八零后或九零后，因此将该变量的缺失值替换为 0。在数据中存在个人年收入高于家庭年收入、家庭可支配年收入高于家庭年收入的个体，由于个体数量较少，因此本文直接将其剔除，最终得到了干净的数据。

在进行描述性分析时，本文将离散型变量进行分类，将工作年限低于 10 年的作为低工作年限，10 年到 20 年之间的为中等工作年限，高于 20 年的为高工作年限。对于驾龄变量，将驾龄低于 10 年的作为低驾龄，10 年到 20 年之间的为中等驾龄，高于 20 年的为高驾龄。对于年度家庭可支配收入变量，低于 10 万元的为低收入家庭，10 万元到 20 万元之间的为中低收入家庭，20 万元到 30 万元之间的为中等收入家庭，30 万元到 50 万元之间的为中高收入家庭，50 万元以上为高收入家庭。

（二）用户特征描述性分析

1、目标客户收入差距分析

目标客户的学历水平主要由七部分组成，由于未受过正式教育和小学的目标客户人

数为 0，因此只统计文化程度从初中到双学位的客户。不同学历个人收入差距的折线图如下图所示，由图 1 可知，个体收入差距大致呈正态分布，其中持有本科学位的客户收入最高，大专学历的客户次之，初中学历的客户收入最低，拥有双学位的客户收入并不高，这与学历和收入呈正相关的说法不符，原因可能在于双学位的人在整个目标客户中人数较少，故虽然该群体收入高，但由于数量少因而在加总收入时，总量较低。总体来看，拥有本科和大专学历的客户收入较为客观，故若提高目标客户的购买意愿，可从这两部分群体入手。

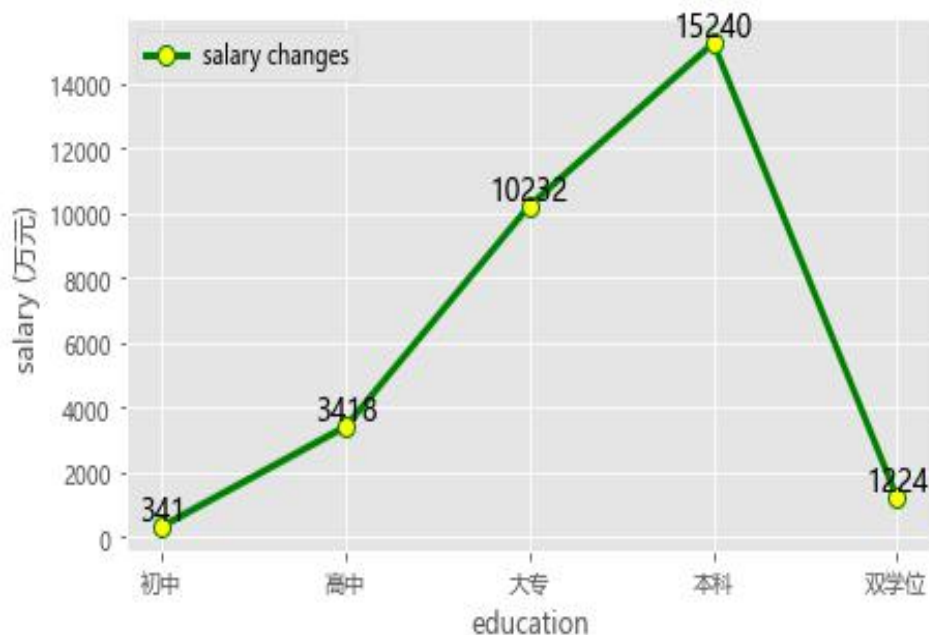


图 1 不同学历个人收入差距

图 2 展示了不同职业的目标客户收入差距的柱形图。由图可以看出，年收入水平最高的是中层管理者，中级技术人员次之，排名第三位的是个体户，收入最低的是其他职业、自由职业以及高层管理者，虽然高层管理者的收入较高，但在给定的目标客户中高层管理者的人数较少，故在加总收入时，总量较低。虽然各类职业的目标客户收入水平仅供参考，但下图可以在一定程度上反映，在购买电动汽车的目标客户中，中层管理者以及中层技术人员、个体户的数量是最多的，因此若要提高客户的购买意愿，可以着重针对这三类人群制定适合其的销售方案，以增加销量。

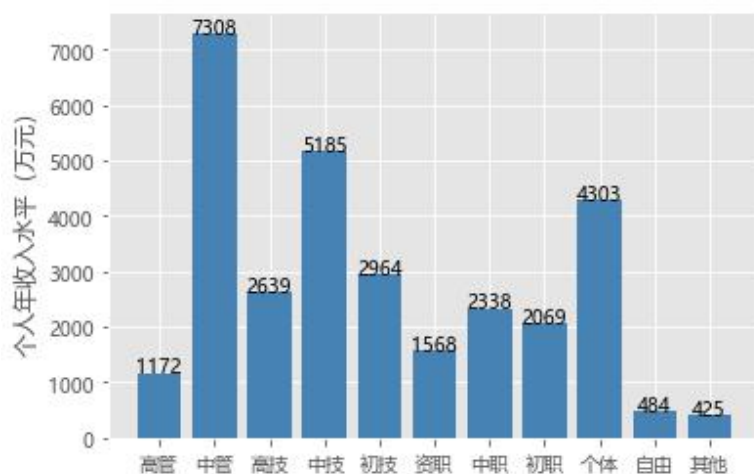


图 2 不同职业个人收入差距

2、目标客户购买意愿分析

不同学历群体对电动汽车这一新兴的事物领域了解深度不同，所以对购买电动汽车是意愿会出现不同想法，本次调研将目标客户的学历主要分为五个水平，分别为双学位/研究生及以上、本科、大专、高中/中专/技校、初中，由图 3 可以直观看出，整体目标客户倾向于不购买电动汽车；在愿意购买的客户中本科学历和大专学历的客户占比很大，并且本科学历客户在整个目标客户中占比最大，他在购买群体和不购买群体的人数都是最多的，其次是大专学历的客户；所以可以主要从本科和大专学历的客户入手来提高客户购买意愿；较高学历双学位/研究生及以上和较低学历初中的客户在整个客户人群中占比较少，且整体偏向于不愿意购买，从这一群体入手来分析提高购买电动汽车意愿也是可行的。

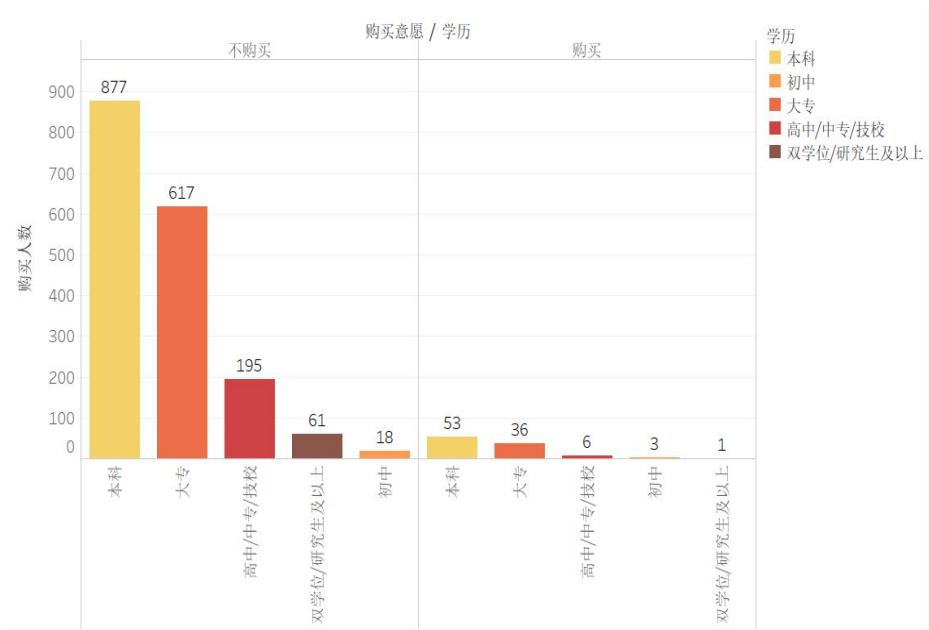


图 3 不同学历群体购买意愿

将家庭可支配收入划分为 50-80（万元）为高收入、30-50（万元）为中高收入、20-30（万元）为中等收入、10-20（万元）为中低收入、0-10（万元）为低收入，在图 4 中蓝色代表目标客户不愿意购买电动汽车的人数，黄色代表愿意购买电动汽车的人数，并且圆圈越大代表客户的人数越多，圆圈越小代表人数越少，由图可以直观看出，蓝色的圆圈大小明显大于黄色，即目标客户倾向于不愿意购买电动汽车，也就是电动汽车的市场前景广阔；在不愿意够买的人群中，中低收入的人群不愿意购买的人数最多，其次是低收入、中等收入、中高收入，高收入人群最少；愿意购买的人群中仍然是中低收入的客户人数最多，其次是中等收入、中高收入、低收入人群，所以可以着重研究中低收入群体来增加购买电动汽车的人数。

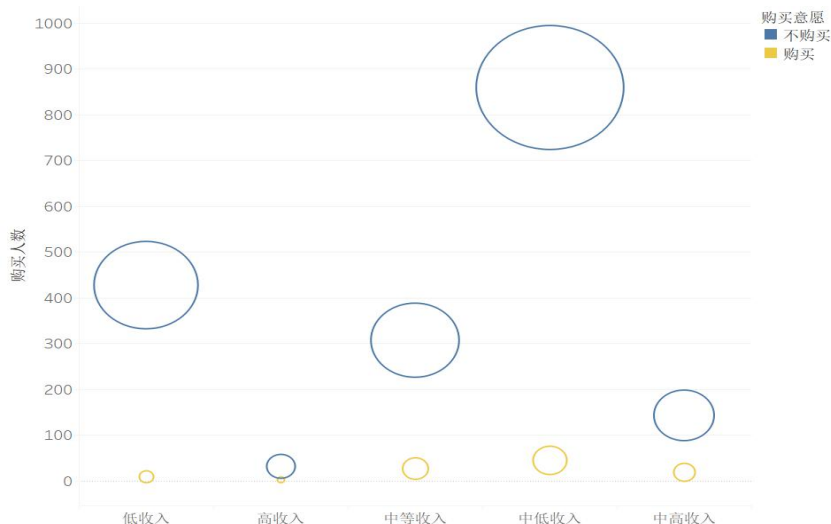


图 4 各收入阶层购买意愿

目标客户的驾龄主要是在 0~30 年之间，根据实际情况将驾龄分为三种情况，其中驾龄在 10 年内为低驾龄、在 10 到 20 年为中等驾龄，在 20 到 30 年的为高驾龄。由图 5 可以看出低驾龄人数最多，其次是中等驾龄，最后为高驾龄；36 个高驾龄的客户中只有三个愿意购买电动汽车，低驾龄中有 1199 个不愿意购买，71 愿意购买，中等驾龄中有 236 人不愿意购买，25 人愿意购买，即中等驾龄的客户中有接近于 10% 的人愿意购买电动汽车，这一比例远大于低驾龄和高驾龄人群，中等驾龄这一部分群体的研究对增加购买电动汽车人数，制定目标客户销售策略至关重要。

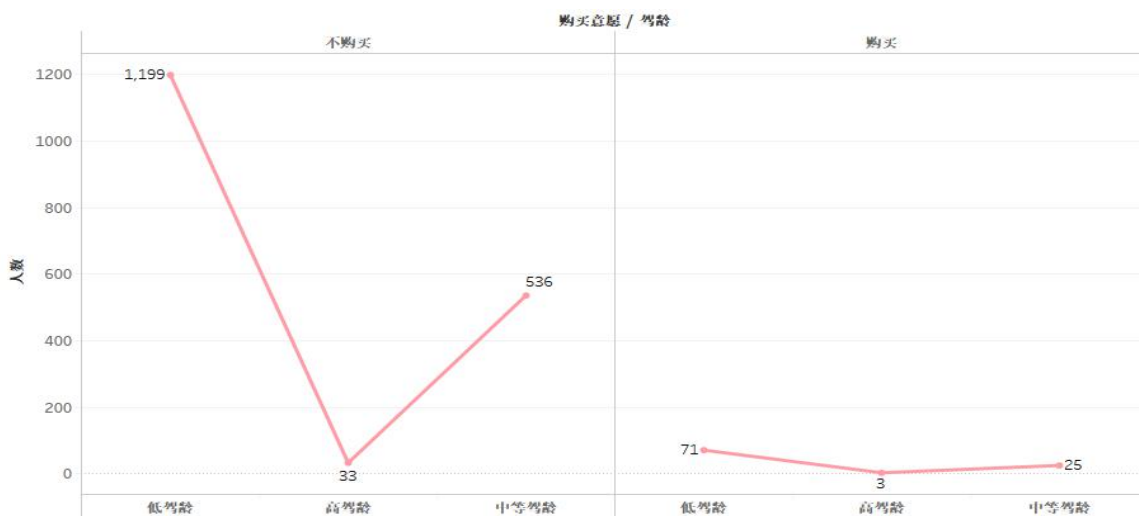


图 5 不同驾龄购买意愿

每个目标客户所处的职位是不同的，对电动汽车在一些领域存在的疑虑也是不同的，这会影响他们购买电动汽车的意愿，图 6 直观展示了不同职位的目标客户对电动汽车的购买意愿情况，可以看到各个职位愿意购买电动汽车的人数分布较为平均，而在不愿意购买的人群中，中层管理者、中级技术人员不愿意购买的人数最多，自由职业者和其他情况不愿意购买的人数最少，这可能是在所选体验的目标客户中自由职业和其他职业的人数少，高层管理者/企业主/老板这部分人群不愿意购买的人数也较少，其他的职位情况不愿意购买的人数大多集中在 100 到 200 人之间，为了提高客户的购买意愿可以在不同职位制定不同销售策略。

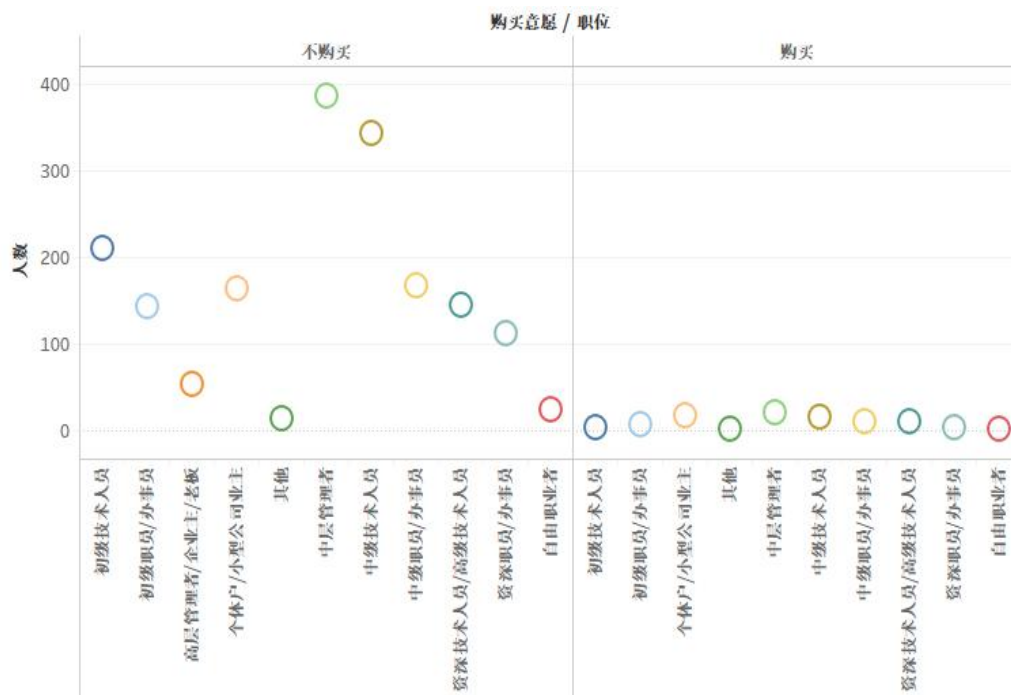


图 6 不同职位购买意愿

(三) 目标客户满意度比较分析

目标客户对汽车公司三款品牌包括合资品牌、自主品牌、新势力品牌电动汽车进行体验，并对不同品牌在电池性能技术、舒适性、经济性、安全性、动力性、驾驶操控性、外观内饰、配置与质量品质这八个维度的满意度进行打分得分，以下雷达图直观展示了客户对三种品牌在八个维度的满意度情况。

图 7 中绿色、蓝色、红色的线分别为为合资品牌、自主品牌、新势力品牌的满意度情况，合资品牌在配置与质量品质、外观内饰、动力性、舒适性这几个方面的满意度是远大于自主品牌、新势力品牌，电池性能、驾驶操控性安全性、经济性这几方面的满意度得分与自主品牌一样，且大于新势力品牌；自主品牌在电池性能技术、舒适性、经济性、安全性、动力性、驾驶操控性的满意度得分大于新势力品牌，在外观内饰、配置与质量品质的满意度得分等于新势力品牌，整体来说合资品牌的满意度最高，其次为自主品牌，最后为新势力品牌，对不同品牌汽车满意度的比较分析可以更好对不同品牌制定适合自身情况的销售策略。



图 7 不同品牌汽车客户满意度比较

三、主成分分析（PCA）

PCA（主成分分析）是应用最广的无监督算法之一，它在尽可能保留原始数据信息的基础上对研究选取的指标数据进行简化，从而达到降维的目的，是一种非常基础的降维算法。由于决定目标客户是否购买电动车的影响因素有很多，有电动汽车本身的因素，也有目标客户个人特征的因素，如下表 1 为本文选取的影响不同品牌电动汽车的销售的因素的指标。为了研究哪些因素可能会对不同品牌电动汽车的销售有影响，首先将所有的影响因素进行降维，然后根据降维后提取出的主成分来分析影响不同品牌电动汽车的销售的因素。

（一）主成分分析适用性检验

由于各指标的单位不一样，所以在进行主成分分析之前先对数据进行标准化，然后对标准化后的数据采用 KMO 试验和 Bartlett 球形试验进行适用性检验，判断 Sig 值与 KMO 值，由表 2 可知，由于数据中 KMO 值=0.863>0.5 表示在容许范围内，因此对购买电动车的影响因素进行主成分分析效果较好，Bartlett 球形检测结果中的近似卡统计值为 31937.65，显著水平 P 值=0.000<0.05，所以可以认为数据中各个影响因素的指标相互独立，由此检验结果知此次目标客户购买电动车的影响因素的数据适合做主成分分析。

表 1 指标含义

指标含义	
A1	体验数据有电池技术性能（电池耐用和充电方便）满意度得分
A2	舒适性（环保与空间座椅）整体表现满意度得分
A3	经济性（耗能与保值率）整体满意度得分
A4	安全性表现（刹车和行车视野）整体满意度得分

A5	动力性表现（爬坡和加速）整体满意度得分
A6	驾驶操控性表现（转弯和高速的稳定性）整体满意度得分
A7	外观内饰整体表现满意度得分
A8	配置与质量品质整体满意度得分
B1	户口情况
B2	本城市居住年数
B3	居住区域
B4	驾龄
B5	家庭成员人数
B6	婚姻情况
B7	孩子个数
B8	出生年
B9	最高学历
B10	工作年限
B11	所在单位的性质
B12	职位
B13	家庭年收入
B14	个人年收入
B15	家庭的可支配年收入
B16	全年房贷的支出占家庭年总收入的比例
B17	全年车贷的支出占家庭年总收入的比例
will	购买意愿

表 2 KMO 和 Bartlett 检验

取样足够的 KMO 度量		0.863
近似卡方		31937.65
Bartlett 的球形度检验	df	300
	Sig.	0.000

（二）求解主成分

表 3 主成分分析结果

成份	解释的总方差								
	合计	初始特征值 a		提取平方和载入			旋转平方和载入		
		方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	6.596	26.383	26.383	6.596	26.383	26.383	6.43	25.72	25.72
2	3.335	13.34	39.723	3.335	13.34	39.723	2.701	10.804	36.524
3	2.888	11.554	51.277	2.888	11.554	51.277	2.545	10.179	46.703
4	1.598	6.391	57.669	1.598	6.391	57.669	2.287	9.148	55.851
5	1.474	5.895	63.564	1.474	5.895	63.564	1.446	5.784	61.635
6	1.27	5.081	68.645	1.27	5.081	68.645	1.444	5.775	67.411
7	1.065	4.258	72.903	1.065	4.258	72.903	1.183	4.733	72.143
8	0.956	3.826	76.728	0.956	3.826	76.728	1.035	4.14	76.283
9	0.9	3.601	80.33	0.9	3.601	80.33	1.012	4.046	80.33

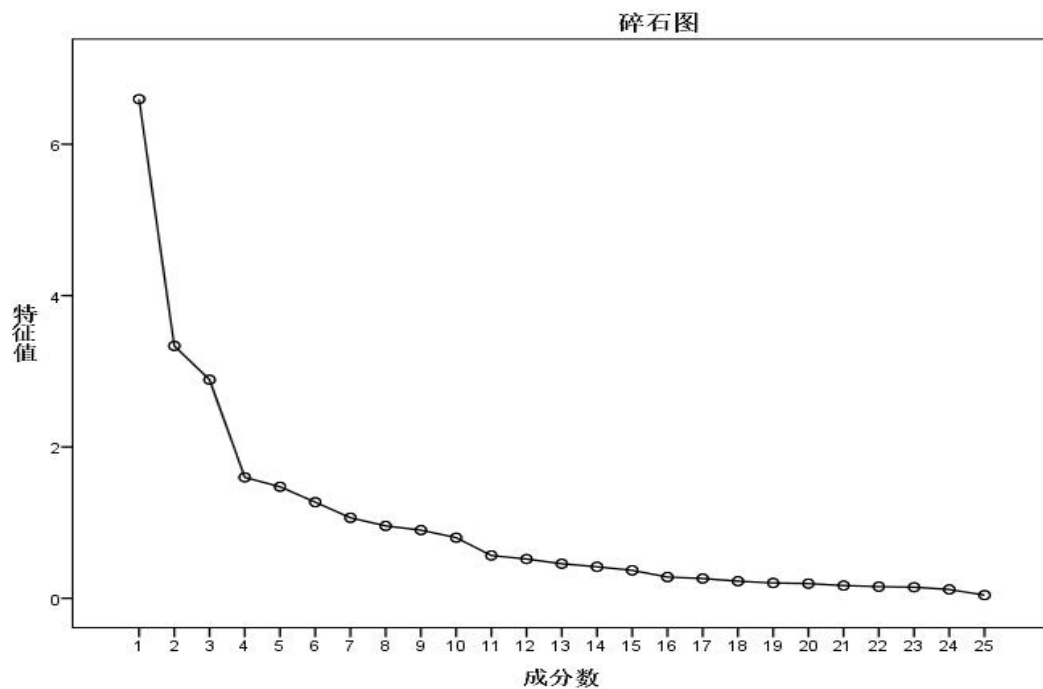


图 8 碎石图

由表 3 和碎石图可以得到前 9 个主成了解释了全部方差的 80.33%，说明提取 9 个主成分能够提取原来 25 个影响购买电动车的因素信息的 80.33%，所以用提取的主成分来研究对不同品牌电动汽车的销售的影响因素是有一定把握的，同时也起到了降维的作用，因此提取 9 个主成分分别为 F1、F2、F3、F4、F5、F6、F7、F8、F9。

F1 的方差贡献率为 26.383%，主要和 A1-A8 有关，即电动汽车本身的因素；F2 的方差贡献率为 13.34%，主要和 B13、B14、B15 有关，即目标客户的收入情况；F3 的方差贡献率为 11.554%，主要和 B10、B8、B4 有关，即目标客户的经历；F4 的方差贡献率为 6.391%，主要和 B5、B6、B7 有关，即目标客户的家庭情况；F5 的方差贡献率为 5.895%，主要和 B1、B2 有关，即目标客户的户口情况；F6 的方差贡献率为 5.081%，主要和 B11、B12 有关，即目标客户的职业情况；F7 的方差贡献率为 4.258%，主要和 B16、B17 有关，即目标客户的贷款情况；F8 的方差贡献率为 3.826%，主要为 B3，即目标客户的居住区

域；F9 的方差贡献率为 3.601%，主要为 B9，即目标客户的学历。

由此可以得到电动汽车本身的因素在电动汽车销售的影响因素中贡献最大，其次是目标客户的收入情况，接着为目标客户的经历，即驾龄、工作年限等因素一定程度上影响电动汽车的销量影响因素，目标客户的家庭情况、户口情况、职业情况的方差贡献率 5%-6%，对购买电动汽车的销量影响因素具有一定的影响，但是小于目标客户经历的影响，目标客户的贷款情况、居住区域、学历的方差贡献率均小于 5%，在一定程度上会影响电动汽车的销量影响因素，但是影响程度较小。

（三）计算主成分得分

根据成分得分系数矩阵可求解得各主成分得分，具体计算过程见附录。

由于有 99 名目标客户购买了体验的电动汽车,为了分析哪些因素可能会对不同品牌电动汽车的销售有影响，对这 99 名购买了电动汽车的客户进行个案汇总，汇总表如下，表中数值均为平均成分得分。

表4 汇总表

品牌类型	F1	F2	F3	F4	F5	F6	F7	F8	F9
合资品牌	0.92607	0.29221	00.12374	-0.21262	0.21389	0.13763	-1.17471	-0.06915	0.13398
自主品牌	1.01787	0.28968	-0.19973	-0.04222	0.19947	0.30960	-1.08910	-0.08784	-0.21924
新势力品牌	1.13394	0.34835	-0.18000	-0.02406	0.07217	-1.10558	-0.64240	-0.00991	0.22988

由汇总表可知 F1 不论是在合资品牌还是在自主品牌、新势力品牌都是影响电动汽车的销售的最重要因素，即电池技术性能（电池耐用和充电方便）、舒适性（环保与空间座椅）、经济性（耗能与保值率）、安全性表现（刹车和行车视野）、动力性表现（爬坡和加速）、驾驶操控性表现（转弯和高速的稳定性）、外观内饰、配置与质量品质这些电动汽车本身的因素主要影响电动汽车的销售，且电动汽车本身的因素对新势力品牌购买的影响要大于自主品牌和合资品牌，所以市场销售在制定相应的销售策略时要首先考虑自身电动汽车的条件。

F2-F9 这些与目标客户个人特征有关的成分对合资品牌、自主品牌、新势力品牌都有一定的影响,但是具体在各个品牌之间的影响也是有区别的,F2 目标客户的收入情况、F4 目标客户的家庭情况、F5 目标客户的户口情况、F7 目标客户的贷款情况、F8 目标客户的居住区域这几个主成分对三个品牌的汽车的影响方向是一致的；F3 目标客户的经历，即驾龄、工作年限等因素对合资品牌的影响是正向的，但是对自主品牌和新势力品牌的影响是负向的；F6 目标客户的职业情况、包括目标客户的职位、所在单位的性质，该因子对合资品牌和自主品牌的销量为正面影响，但是对新势力品牌的销量的影响却是负面的；F9 目标客户的最高学历对自主品牌有负面影响，在合资品牌和新势力品牌是正面影响，因此市场销售在制定不同品牌销售策略时要对不同特征的目标客户具体定制不同的销售策略，实施精准营销。

四、神经网络模型

（一）多层感知器神经网络模型（MLP）

1、多层感知器神经网络模型（MLP）

人工神经网络（Artificial Neural Networks, ANN）是一种可用于处理具有多个节点和多个输出点的实际问题的网络结构。它从信息处理角度对人脑神经元网络进行抽象，建立某种简单模型，按不同的连接方式组成不同的网络。它由大量的节点（或称神经元）之间相互联接构成。每个节点代表一种特定的输出函数，称为激励函数（activation function）。每两个节点间的连接都代表一个对于通过该连接信号的加权值，称之为权重，这相当于人工神经网络的记忆。网络的输出则依网络的连接方式、权重值和激励函数的不同而不同。网络自身通常都是对自然界某种算法或者函数的逼近，也可能是对一种逻辑策略的表达。

多层感知器（Multilayer Perceptron, MLP）又称为多层前馈神经网络，具有出色的非线性映射能力、较高的并行性以及全局优化的特点，如今在图像处理、预测系统、模式识别等方面取得了显著成果。多层感知器神经网络的学习过程是从网络的输入层开始的，经隐含层逐步处理，并传向输出层，每一层神经元的状态只影响下一层神经元的状态，通过对各层权值进行修改，使误差最小。在 Keras 的深度学习框架下存在两种类型的模型，序贯模型（Sequential）和函数式模型（Model），本文利用 python，在 Keras 的深度学习框架下使用多层感知器神经网络来实现对二分类变量消费者购买意愿的预测。在模型选取时采用序贯模型进行神经网络的构建，序贯模型具有单输入单输出的特征，一条路通到底，层与层之间只有相邻关系，没有跨层连接，因此这种模型编译速度快，普适性强。

2、神经网络的构建

本文利用全部特征变量：汽车体验满意度、目标客户个人特征和结果变量判断目标客户的购买行为，其中结果变量为购买意愿，数值 1 代表有购买行为，0 代表没有购买行为。对清洗后的数据进行分析，建立多层感知器神经网络模型，并预测部分用户的购买行为。在模型构建中，本文建立序贯模型的 MLP 二分类预测模型，将训练集设置为样本量的 90%，将测试集设置为样本量的 10%，样本总量为 1867，训练集数据量为 1680，测试集数据量为 187。

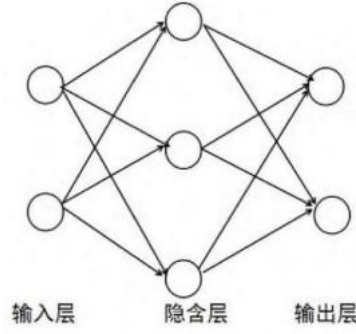


图 9 MLP 神经网络基本结构图

MLP 网络模型算法核心思想是：通过前向通路（图 8 箭头的方向）得到误差，再把该误差反向传播实现权值 w 的修正。MLP 模型网络初始权值是任意设置的，对每个输入样本 p ，网络输出与期望输出的平方误差为：

$$E = \sum_{j=1}^j (t_j - y_j^k)^2 \quad (1)$$

式中， t_j 为第 p 个输入样本输出单元 j 的期望输出， y_i 为第 p 个输入样本的网络输出。式(1) 中 y_j^k 是由上一层（即第 $k-1$ 层）所有神经元的输出经加权激活后得到，而第 $k-1$ 层神经元的输出又由第 $k-2$ 层的所有神经元的输出经加权激活后得到，因此可以说误差 E 是全体权值 w 的函数，通过改变权值 w ，就可达到使误差 E 最小。即误差大，权值的变化就越大，而当权 值改变时，误差就要重新计算。这样两者相互作用，不断迭 代，直到误差小于某个值（即收敛）为止。[5-6] 为使式(1)中的 E 值最小，本文采用的梯度下降法改变权值，并引入学习速率 η 及常数参数 α ，避免改变 η 值。选取不好带来的问题，同时使误差 E 达到最小。则变化率公式为：

$$\Delta w(t) = -\eta \frac{dE}{dw}(t) + \alpha \Delta w(t-1) \quad (2)$$

式中， t 表示当前， $t-1$ 表示上一次， $t+1$ 表示下一次。

因此本次的 w 变化率不仅与 E 的导数有关，还与上次 w 的变化率有关。则其权值修正公式为：

$$w(t+1) = w(t) + \Delta w(t) \quad (3)$$

3、神经网络结果分析

选取汽车体验满意度、目标客户个人特征作为特征输入并进行归一化处理，特征维度为 25。本文采用序贯模型，输入层的特征维度设置为 25，迭代次数设置为 15 次，每次训练样本为 2520。损失函数用使用交叉熵指标评价，学习率取 0.001，对测试集的目标客户购买行为进行预测。

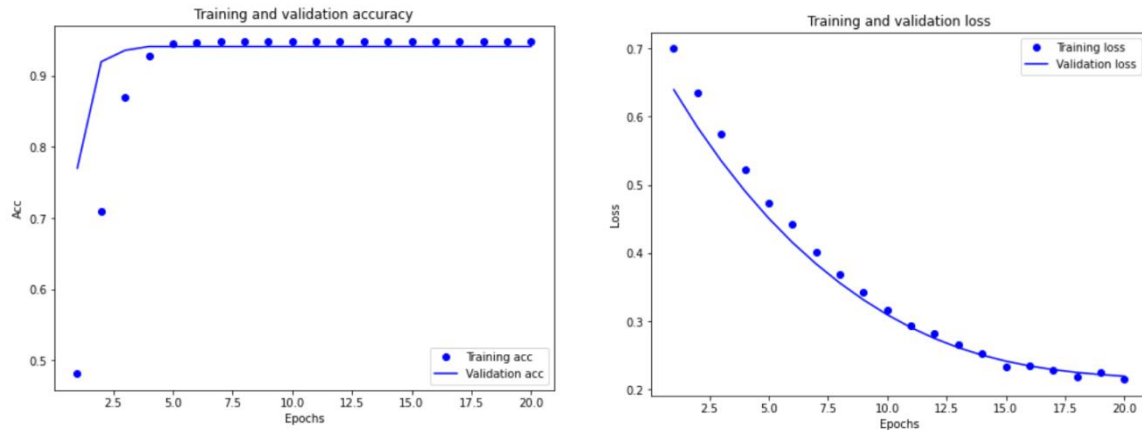


图 10 模型评价 acc_loss 图

模型评价曲线如图 10 所示，由左图可以看出，模型迭代第四次后倾向于收敛，没有太大震荡。最终收敛的预测准确率为 0.948，准确率高的原因是由于发生购买行为的总人数为 99，大部分用户并未发生购买行为。综上所述，模型的预测准确度较高。右图显示了模型在训练时的 loss 曲线收敛情况，图中的 20 为神经网络模型设置的 epoch，即训练多少轮。根据 loss 曲线可以看出，该模型最后的 loss 稳定在 0.2。

利用 MLP 模型的训练结果对待判定的数据进行预测，目标客户预测结果全部为不购买。在对模型结果进行评估时，考虑到本文所利用的数据样本量较小，在进行训练时会有较大误差，并且本文利用序贯模型，仅存在前向传播而不存在层与层之间的反向传播，因此该方法的预测效果并不理想，因此本文采用附加动量的 BP 神经网络做进一步分析。

（二）BP 神经网络模型

1、BP 神经网络模型

BP 神经网络是一种多层前馈网络，结构主要由 3 部分组成：输入层、隐含层和输出层。三层网络结构是常用的结构，其拓扑结构图如图 11 所示。

人工神经网络无需事先确定输入输出之间映射关系的数学方程，仅通过自身的训练，学习某种规则，在给定输入值时得到最接近期望输出值的结果。作为一种智能信息处理系统，人工神经网络实现其功能的核心是算法。BP 神经网络是一种按误差反向传播（简称误差反传）训练的多层前馈网络，其算法称为 BP 算法，它的基本思想是梯度下降法，利用梯度搜索技术，以期使网络的实际输出值和期望输出值的误差均方差为最小。

基本 BP 算法包括信号的前向传播和误差的反向传播两个过程。即计算误差输出时按从输入到输出的方向进行，而调整权值和阈值则从输出到输入的方向进行。正向传播时，输入信号通过隐含层作用于输出节点，经过非线性变换，产生输出信号，若实际输出与

期望输出不相符，则转入误差的反向传播过程。误差反传是将输出误差通过隐含层向输入层逐层反传，并将误差分摊给各层所有单元，以从各层获得的误差信号作为调整各单元权值的依据。通过调整输入节点与隐层节点的联接强度和隐层节点与输出节点的联接强度以及阈值，使误差沿梯度方向下降，经过反复学习训练，确定与最小误差相对应的网络参数(权值和阈值)，训练即告停止。此时经过训练的神经网络即能对类似样本的输入信息，自行处理输出误差最小的经过非线性转换的信息。

BP 学习算法的一般权重调整如式 (4) 所示

$$\begin{aligned}\Delta \omega_{jk} &= -\eta \frac{\partial E}{\partial \omega_{jk}} = \eta \delta_k y_j = \eta (d_o(n) - y_o(n)) y_j f(\text{net}_k) \\ \Delta \omega_{ij} &= -\eta \frac{\partial E}{\partial \omega_{ij}} = \eta \delta_j x_i = \eta \left(\sum_{k=1}^L \delta_k \omega_{jk} \right) x_i f(\text{net}_j)\end{aligned}\quad (4)$$

其中, ω_{jk} 是输出层与隐含层的网络权重, ω_{ij} 是输入层与隐含层的网络权重, $\Delta \omega_{jk}$ 与 $\Delta \omega_{ij}$ 是权重增量, x_i 是第 i 个节点输入, y_j 是第 j 个节点输出, E 是神经网络输出节点的误差函数, 即图 11 BP 神经网络拓扑图

$$E = \frac{1}{2} \sum_{o=1}^q (d_o(n) - y_o(n))^2 \quad (5)$$

δ_k 和 δ_j 是从外层到内层的误差反向传播信号, L 是输出神经元的数量, $d_o(n)$ 是网络的期望输出值, $y_o(n)$ 是网络实际输出, $f(\text{net}_k)'$ 和 $f(\text{net}_j)'$ 是输出和隐含层的传递函数的导数, 负号表示梯度下降, 常数 $\eta \in (0, 1)$ 反映网络学习率。

虽然反向传播算法是人工神经网络研究的一个重要进展, 但基本 BP 算法的训练速度太慢, 且多层神经网络的性能表面可能具有许多局部最小值, 这使得难以收敛到全局最小点。因此, 必须改进传统的 BP 算法进行客户购买意愿的识别。

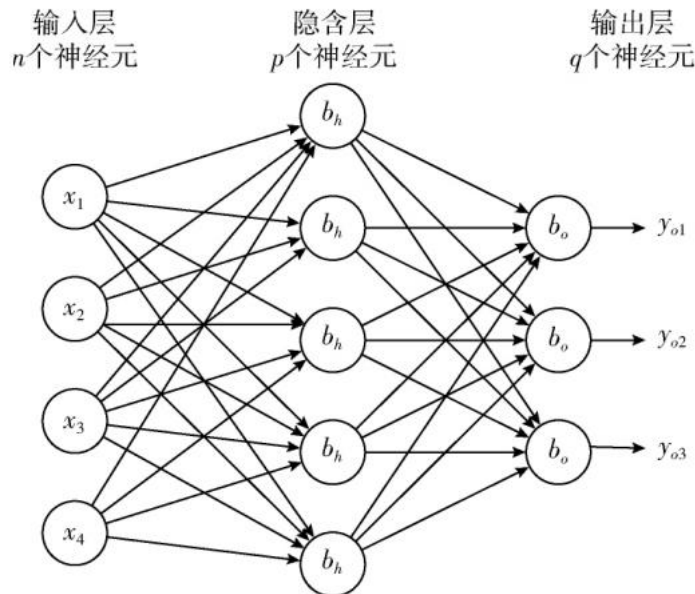


图 11 BP 神经网络拓扑图

2、附加动量的 BP 神经网络的构建

为了提高 BP 神经网络对目标客户购买行为识别的准确率，本文利用附加动量法，降低在 BP 神经网络在训练过程中陷入局部极小值的概率，提高收敛速度。附加动量法是将动量因子与本次误差计算权重变化量两者的叠加和，作为本次的实际权重变化量进行继承传递，权值调节公式为

$$\Delta\omega(n+1) = \beta(\omega(n+1) - \omega(n) + \alpha(1 - \beta)[\partial E(n)/\partial \omega(n)]) \quad (6)$$

其中， β 为动量因子， α 为学习速率， n 为训练次数， $\partial E(n)/\partial \omega(n)$ 为权值 $\omega(n)$ 的梯度。当模型误差减小时，我们可以在不同的训练周期中采用不同的动量因子来提高网络性能。在误差表面的平坦部分中，增加动量因子有助于重量矢量逃离误差表面的平坦区域，这可以加速收敛。相反，在误差表面的陡峭部分，减小动量因子避免了网络的不稳定性，可以根据关于权向量的误差函数的梯度来自适应地调节动量因子。即

$$\beta = \begin{cases} 0, & E[\omega(n+1)] \geq pE[\omega(n)] \\ e^{-\lambda - \|\frac{\partial E}{\partial \omega}\|}, & E[\omega(n+1)] < pE[\omega(n)] \\ \beta, & \text{其他} \end{cases} \quad (7)$$

其中， p 是最大允许误差变化率， λ 是随机数且 $0 < \lambda < 1$ 。

当 BP 神经网络权值进入误差曲面底部的平坦区域时， $\partial E(n)/\partial \omega(n)$ 将不断变小，直至有

$$\Delta\omega(n+1) = \beta(\omega(n+1) - \omega(n)) = \Delta\omega(n) \quad (8)$$

从而避免了 $\Delta\omega(n) = 0$ ，有利于网络跳出误差曲面局部最小值。标准 BP 算法收敛缓慢的另一个重要原因是学习率在训练过程中是不变的。学习率影响收敛速度，在迭代过程中进行学习率自适应调整，确保 BP 神经网络始终具有最大可接受的训练学习率。即

$$\eta(n) = \begin{cases} 1.05\alpha(n-1), & E(n) < E(n-1) \\ \alpha(n-1)e^{-\rho}, & E(n) > E(n-1) \end{cases} \quad (9)$$

其中， $\rho \in [0.0001, 0.001]$ ，当网络误差处于下降状态且下降趋势明显时，则该方法的学习率可以按照一定比例增大。由于附加动量法可以发现更好的全局解，而自适应学习率法可以减少 BP 算法的训练次数，可以将上述两种方法结合起来，充分发挥其优势^[1]。改进 BP 神经网络识别过程如图 12 所示。

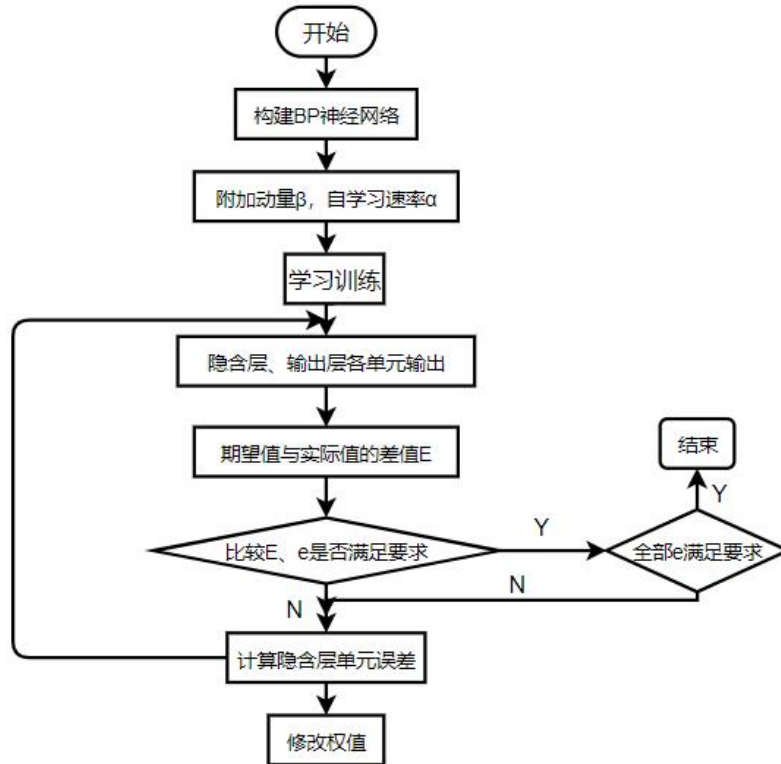


图 12 改进 BP 神经网络识别流程

许多神经网络往往都是设定好层数，然而 BP 神经网络能够按照实际情况来确定其层数，并且可以根据输入层和输出层的维度来决定 BP 神经网络隐含层的层数^[2]。本文建立一个包含三个隐含层的附加动量的 BP 神经网络模型，输入层的节点数目为输入矢量的维数，即 26 维。输出层通常代表系统需要实现的功能目标，本文旨在分析目标客户满意度情况、用户特征和用户购买情况之间的关系，来构建目标客户购买电动车可能性的预测模型，以实现精准预测。因此本研究的输出节点是目标客户的购买意愿，即输出层节点有 1 个。

本文利用 MATLAB2012a 在设定隐含层节点数目时，Funahash^[3]认为当 BP 神经网络隐含层节点的函数是单调递增的连续函数时，只有一个隐含层的前馈神经网络可以实现以任何精确度来逼近任意线性函数。Hornik^{[4][5]}和陈天平^[6]已经验证隐含层节点函数是有界函数是必要条件，而要求是单调递增是非必要条件。因此基于只包含一层隐含层的神经网络，其隐含层节点函数是 Sigmoid，是足以实现网络逼近。陈天平^[7]还提及到，之所以出现过度拟合的现象是因为 BP 神经网络的隐含层节点数目设置太多，从而会降低 BP 神经网络的泛化能力。董聪^[8]认为在精确度符合要求的条件下，将逼近函数的阶数设置越少，那么效果会更好，低阶数的逼近函数往往能够很好地避免过度拟合的出现，从而提高 BP 神经网络的泛化能力。换言之，当精确度符合要求的前提下，BP 神经网络隐含层节点数目设置越少效果就越好。本研究通过对目标客户购买可能性进行不断训练与测试，最终得出该模型隐含层节点数目设置为 25 个为最佳。此外，设定学习率为

0.01，续联函数为 logsig 函数。

3、BP 神经网络结果分析

将电池技术性能：舒适性（环保与空间座椅）、经济性（耗能与保值率）、安全性表现（刹车和行车视野）、动力性表现（爬坡和加速）、驾驶操控性表现（转弯和高速的稳定性）、外观内饰整体表现、配置与质量品质以及目标客户体验者个人特征的信息作为特征输入后，对合资品牌、自主品牌和新势力品牌三种电动汽车建立客户挖掘模型，由于附加动量的 BP 神经网络可以减小训练次数，且在相同训练次数内，训练误差下降速度快，因此对目标客户购买意愿进行判断时，准确性较高。

表 5 展示了在训练大量样本之后，深入挖掘客户对不同品牌电动汽车的满意度评价以及个人特征，最后将 15 个目标客户样本作为测试集，预测出他们的购买意愿，如下表所示。为方便阅读，将系统出的结果中的“1”替换为“否”，将系统结果“2”替换为“是”，从而得到 15 名目标客户中，仅客户 3、客户 8 以及客户 11 三位愿意购买电动汽车，其余客户均不愿购买。在意愿购买的三位客户中，客户 3 愿意购买合资品牌汽车，客户 8 愿意购买自主品牌汽车，客户 11 愿意购买新势力品牌汽车。

表 5 目标客户购买可能性

客户编号	品牌编号	是否会购买?
1	1	否
2	1	否
3	1	是
4	1	否
5	1	否
6	2	否
7	2	否
8	2	是
9	2	否
10	2	否
11	3	是
12	3	否
13	3	否
14	3	否
15	3	否

五、目标客户销售策略

由于提高体验满意度 5%的服务难度是提高体验满意度 1%服务难度的 5 倍，本文分别就附件 3 每个品牌中未购买电动汽车的目标客户，抽取客户实施销售策略。由于多层感知器神经网络模型的结果中，测试集中没有目标客户愿意购买电动汽车，这与训练集中的结果出入较大，因此本文利用附加动量的 BP 神经网络模型测试集结果进行分析。分别抽取合资品牌中不愿购买的客户 1、自主品牌汽车中不愿购买的客户 6 以及新势力品牌汽车中不愿购买的客户 15 进行销售策略分析。

对于客户编号为 1 的客户，从附件 3 中提供的满意度信息来看，他对合资品牌电动汽车的外观内饰整体表现以及配置与质量品质的满意度最高，达到 99.9，可见他对汽车的美观程度体验感较好，是追求美观与品质的客户，而对舒适性（环保与空间座椅）整

体表现、经济性（耗能与保值率）、安全性表现（刹车和行车视野）以及动力性表现（爬坡和加速）评价不高，可见这些因素是阻碍他购买合资品牌汽车的重要因素。因此在对客户 1 进行销售时，应着重强调客户 1 满意的部分，如汽车的美观程度以及配置的质量，首先给客户营造一种较高的体验感，其次，针对客户满意度较差的舒适性、安全性、经济性等，可为客户进行深入的讲解，并邀请客户进行现场体验，来提高自身的服务力度，从而增强客户的满意度。

针对客户 6，可以看出他对自主品牌汽车的满意度中，电池技术性能（电池耐用和充电方便）的满意度是最高的，配置与质量品质次之，动力性表现（爬坡和加速）最差，达到了 79.54，总体而言，客户 6 的满意度远远低于客户 1 对合资品牌汽车的满意度。因此，在提供销售建议方面，首先应从客户满意度较高的电池技术性能入手，进而推荐自主品牌汽车的配置与质量品质，给客户留下良好的第一印象，进而，针对满意度较低的舒适性、经济性、安全性方面，可以邀请客户进行试车体验，在体验的过程中充分展示自主品牌汽车的优势，给客户一种极佳的体验感，最后，针对客户体验感最差的动力性如爬坡和加速，着重为客户介绍该品牌汽车在这方面的参数以及该汽车的优势，并对比市场上同等价位的汽车，总结出自主品牌汽车的优点。

针对客户 11，他对新势力品牌汽车的整体满意度在这五位客户中较低，尤其是对动力性表现（爬坡和加速）、经济性（耗能与保值率）、外观内饰整体表现、配置与质量品质方面的体验感不容乐观，而对电池技术性能（电池耐用和充电方便）、安全性表现（刹车和行车视野）以及驾驶操控性表现（转弯和高速的稳定性）方便较为满意。因此针对客户 11 进行销售时，也应从满意度高的方面入手，其次诚邀客户进行试车体验，在这一过程中形象的介绍这些方面的优点，让客户有一个更加清晰的感知。

六、结论与建议

（一）开启蓝海市场、针对特定用户精准营销

在本文的描述性分析中，研究得到个人学历与收入大致成正相关，因此该公司可以根据这一特性对高学历的客户，如大专和本科学历群体进行精准营销，提升其购买意愿。在该公司的客户信息中发现，本科与大专学历的人数占比接近 80%，在购买行为分析中，这两类购买者在总购买人数中占比 90%，因此该类客户具有巨大的市场潜力与开拓价值。

在对客户不同职业的收入差距分析中，中层管理者、中级技术人员、个体户的个人年收入水平较高。在不同职位购买意愿分析中，中层管理者与中级技术人员在不购买新能源汽车的总人数中所占比例较高，对比分析可以发现该类群体年收入较高，但目前尚未购买该公司汽车，因此该公司应拓宽销售渠道，从客户的职业角度出发，为之提供针对性营销战略，以在高潜力的新市场中创造收益。

对家庭可支配收入的研究中，家庭可支配收入处于 10 万元到 20 万元之间的中低收入的家庭在购买新能源汽车的人数中所占比例较高，其次是低收入家庭，而高收入家庭与中高收入家庭的购买人数较少。因此该公司可针对收入较高的客户开展新能源汽车的性能宣传与普及，提升这部分人群的购买意愿。其次，研究得到低驾龄群体的购买意愿最强烈，并且这部分人群多为年轻人，对新兴事物的接收度较高。该公司应超越传统市场与传统营销模式，开创出囊括新客户的市场，开拓未知的产业市场与市场空间。

（二）提高电动汽车产品质量、性能和服务

1. 分析可知目标客户在购车时较为关注电动汽车自身配置，因此对于汽车生产企业，应该着眼于汽车性能的提升。由于电动汽车在满足使用需求的同时，还可以实现节能、减排、低碳，所以客户在购买电动汽车时不仅会考虑驾驶汽车时的舒适性，还会考虑到电动汽车的可持续发展，所以对广大汽车生产商和经销商来说，汽车的电池质量、性能和服务至关重要，是其能否在汽车消费市场上占有一席之地之关键因素，具体来说①当前的电动汽车电池长时间使用会使续航里程明显缩减，汽车公司可以加大力度研发技术来提高电池续航能；②城市充电桩配备严重不足，导致电动汽车不宜在高速公路上行驶所以汽车公司需要增加城市充电桩；③客户对电动汽车的安全舒适性有一定的要求，汽车公司的电动汽车生产更具有科技感和未来感以及极佳的起步加速性；④汽车公司应降低汽车的耗能，生产研发低油耗、低噪声、油电切换平稳的电动汽车。

2. 汽车公司要提供渠道给消费者了解有关电动汽车属性的信息，对于广大消费者而言，追求性价比、舒适度和安全性是普遍的电动汽车消费心理，增加消费者对汽车产品的认识与理解，可以帮助他们建立起对于汽车价值的感知，当消费者对汽车产品有良好的感知价值，才能增强消费者满意度。

（三）对不同品牌目标客户实施精准营销。

受社会、家庭、文化、经济等多种因素的影响，不同品牌的电动汽车消费者在购车原因上存在明显差异，市场销售在制定不同品牌销售策略时要对不同特征的目标客户具体定制不同的销售策略，知道目标客户会对哪些产品买单，这样就能对此实施精准营销，全面满足客户需求，降低营销成本。

综合来说电动汽车可以大量在那些家庭收入较高、在本城市居住有一定年限，且家庭成员不多，有较少贷款的客户群体之间推广，合资品牌可以在高驾龄、年纪较小、工作时间较长、学历较高且工作较为稳定的人群中推广；自主品牌适合低驾龄刚工作学历一般的人群中宣传推广；新势力品牌适合低驾龄刚工作学历较高的人群中宣传推广。

参考文献

- [1] 凡甲甲, 祁云嵩, 葛霓琳. 基于改进 BP 神经网络的船舶轨迹识别方法[J]. 计算机工程与设计, 2019, 40(12): 3639-3644.
- [2] 王娟. 基于 BP 神经网络的高原农产品销量预测[D]. 昆明理工大学, 2020.
- [3] Funahshi K I. On the Approximate Realization of Continuous Mappings by Neural Networks[J]. Neural Network, 1989, 2: 183-192.
- [4] Hornik K, Stinchcombe M, White H. Multilayer Feedforward Networks Are Universal Approximators[J]. Neural Networks. 1989, 2: 359-366.
- [5] Hornik K. Approximation Capabilities of Multilayer Feedforward Networks Neural[J]. Neural Network, 1991, 4: 551-557.
- [6] 陈天平. 神经网络及其在系统识别应用中的逼近问题[J]. 中国科学(A 辑), 1994, 24 (1): 1-7.
- [7] 董聪. 前向网络全局最优化问题研究[J]. 中国科学基金, 1997, 1: 23-29.
- [8] 丁一然. 基于神经网络的航空发动机故障诊断[J]. 电子制作, 2021(08): 42-44+78.

附录一 数据清洗与描述性分析代码

1、数据清洗 SQL 语句

```
use test;
#删除 a1、a3、a5、B17 列大于 100 的行
delete from `data` where a1>100;
delete from `data` where a3>100;
delete from `data` where a5>100;
delete from `data` where B17>100;
#将 B7 列的 null 值替换为 0
update `data` set B7=0 where B7 like '%NULL%';
#将 B14>B13 的行剔除
select * from `data` where B14>B13;
delete from `data` where B14>B13;
#将 B15>B13 的行剔除
select * from `data` where B15>B13;
delete from `data` where B15>B13;
```

图 6 代码

```
import matplotlib.pyplot as plt
```

2、图 1 代码

```
x1 = ['初中', '高中', '大专', '本科', '双学位']
# 体重
y1 = [341, 3418, 10232, 15240, 1224]
# 设置画布大小
plt.figure(figsize=(7, 4))
# 数据
plt.plot(x1, y1, label='salary changes', linewidth=3, color='g', marker='o',
         markerfacecolor='yellow', markersize=8)
# 横坐标描述
plt.xlabel('education')
# 纵坐标描述
plt.ylabel('salary (万元)')
# 设置数字标签
for a, b in zip(x1, y1):
    plt.text(a, b, b, ha='center', va='bottom', fontsize=13)
plt.legend()
plt.show()
```


3、图2 代码

```
import matplotlib.pyplot as plt
import pandas as pd
b = pd.read_excel('D://a.xlsx')
plt.style.use('ggplot')
plt.rcParams['font.sans-serif'] = ['Microsoft YaHei']
bar_width = 1000000
plt.bar(x = range(b.shape[0]),
        height = b.salary,
        tick_label = b.position,
        color = 'steelblue',
        )
plt.ylabel('个人年收入水平（万元）')
for x,y in enumerate(b.salary):
    plt.text(x,y+0.1,"%s"%round(y,1),ha='center')
plt.show()
```

4、雷达图 Python 代码

```
import numpy as np
import matplotlib.pyplot as plt

results = [{"电池技术性能": 78, "舒适性": 79, "经济性": 76, "安全性": 79, "动力性": 78, "驾驶操控性": 78, "外观内饰": 79, "配置与质量品质": 78},
           {"电池技术性能": 78, "舒适性": 78, "经济性": 76, "安全性": 79, "动力性": 77, "驾驶操控性": 78, "外观内饰": 78, "配置与质量品质": 77},
           {"电池技术性能": 77, "舒适性": 77, "经济性": 75, "安全性": 78, "动力性": 76, "驾驶操控性": 77, "外观内饰": 78, "配置与质量品质": 77}]
data_length = len(results[0])
# 将极坐标根据数据长度进行等分
angles = np.linspace(0, 2*np.pi, data_length, endpoint=False)
labels = [key for key in results[0].keys()]
score = [[v for v in result.values()] for result in results]
# 使雷达图数据封闭
score_a = np.concatenate((score[0], [score[0][0]]))
score_b = np.concatenate((score[1], [score[1][0]]))
score_c = np.concatenate((score[2], [score[2][0]]))
angles = np.concatenate((angles, [angles[0]]))
labels = np.concatenate((labels, [labels[0]]))
# 设置图形的大小
fig = plt.figure(figsize=(7, 5), dpi=100)
# 新建一个子图
```

```

ax = plt.subplot(111, polar=True)
# 绘制雷达图
ax.plot(angles, score_a, color='g')
ax.plot(angles, score_b, color='b')
ax.plot(angles, score_c, color='r')
# 设置雷达图中每一项的标签显示
ax.set_thetagrids(angles*180/np.pi, labels)
# 设置雷达图的0度起始位置
ax.set_theta_zero_location('N')
# 设置雷达图的坐标刻度范围
ax.set_rlim(75, 80)
# 设置雷达图的坐标值显示角度，相对于起始角度的偏移量
ax.set_rlabel_position(270)
ax.set_title("不同品牌汽车满意度比较")
plt.legend(["合资品牌", "自主品牌", "新势力品牌"], loc='upper right')
plt.show()

```

附录二 各因子得分计算过程

$F1=0.139A1+0.146A2+0.134A3+0.144A4+0.143A5+0.142A6+0.140A7+0.141A8-0.022B1-0.013B2-0.004B3-0.008B4-0.017B5-0.015B6-0.013B7+0.11B8-0.018B9-0.012B10-0.003B11-0.011B12+0.002B13+0.001B14+0.001B15-0.016B16-0.005B17$

$F2=-0.004A1+0.008A2-0.003A3-0.002A4+0.013A5-0.003A6+0.001A7-0.002A8+0.038B1-0.05B2+0.008B3-0.012B4+0.032B5+0.002B6+0.021B7+0.031B8-0.023B9-0.026B10-0.008B11-0.046B12+0.387B13+0.359B14+0.334B15+0.018B16+0.062B17$

$F3=-0.019A1-0.009A2-0.003A3-0.008A4-0.001A5+0.006A6-0.008A7-0.011A8-0.12B1+0.78B2+0.021B3+0.353B4-0.164B5-0.032B6-0.008B7-0.388B8+0.049B9+0.39B10+0.073B11-0.023B12-0.045B13-0.002B14-0.032B15+0.04B16-0.009B17$

$F4=-0.007A1-0.003A2-0.021A3-0.007A4-0.014A5-0.016A6-0.005A7-0.007A8+0.26B1-0.049B2-0.001B3-0.11B4+0.439B5+0.398B6+0.373B7+0.034B8+0.013B9-0.047B10+0.04B11+0.009B12+0.028B13+0.011B14+0.014B15+0.028B16-0.031B17$

$F5=0.015A1-0.018A2-0.014A3-0.026A4-0.007A5-0.017A6-0.021A7-0.017A8+0.659B1+0.474B2-0.047B3-0.018B4+0.072B5-0.043B6-0.046B7+0.071B8-0.034B9-0.065B10-0.005B11+0.066B12+0.036B13-0.016B14-0.004B15-0.176B16+0.190B17$

$F6=-0.023A1+0.002A2-0.004A3-0.009A4+0.011A5-0.002A6-0.012A7+0.002A8+0.048B1+0.01B2+0.005B3+0.034B4+0.045B5+0.025B7-0.024B8+0.048B9+0.018B10+0.618B11+0.574B12-0.044B13+0.008B14-0.037B15+0.093B16-0.07B17$

$F7=-0.025A1-0.004A2-0.013A3-0.009A4-0.003A5-0.006A6-0.014A7-0.003A8+0.132B1-0.121B2+0.001B3+0.032B4+0.032B5+0.007B6-0.036B7-0.003B8-0.012B9+0.014B10-0.01B11+0.035B12+0.166B13+0.085B14-0.107B15+0.616B16+0.654B17$

$F8=0.013A1+0.001A2-0.026A3+0.01A4-0.019A5+0.002A6-0.008A7-0.004A8+0.031B1-0.08B2+0.817B3-0.058B4+0.005B5-0.013B6-0.016B7-0.051B8+0.015B9+0.045B10-0.$

107B11+0.085B12+0.036B13-0.001B14+0.012B15-0.414B16+0.348B17
 F9=-0.046A1-0.008A2-0.031A3-0.019A4-0.023A5+0.012A6+0.009A7+0.01A8-0.04B
 1-0.002B2+0.007B3+0.121B4-0.035B5+0.028B6+0.045B7-0.016B8+0.991B9-0.04B10+0.
 015B11+0.074B12+0.005B13-0.021B14-0.04B15-0.094B16+0.07B17

附录三 神经网络代码

1、Python 多层感知器神经网络模型（MLP）代码

```
#python 导入相应库
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt #2D 绘图库
from sklearn.datasets import load_iris
from sklearn.decomposition import PCA
%matplotlib inline
from sklearn.model_selection import train_test_split
from sklearn import datasets
from sklearn import preprocessing
from keras.preprocessing.image import load_img, img_to_array

import math
import os
from tensorflow.keras import Sequential, optimizers
from tensorflow.keras.layers import Dense, BatchNormalization, Dropout, Activation
from tensorflow.keras.models import load_model
from tensorflow.keras.callbacks import ReduceLROnPlateau
from tensorflow.keras import backend as K
import warnings
warnings.filterwarnings('ignore')

data = pd.read_csv('E:/华数杯/data111.csv')
data.head(7)

from sklearn.model_selection import train_test_split
from sklearn import datasets
```

```

# 10%用于测试集, 90%用于训练集
train, test= train_test_split(data, test_size=0.10, random_state=20)

def get_y_train():
    y_train = np.array(train['will'])

    return y_train

def get_y_test():
    y_test = np.array(test['will'])

    return y_test

def get_X_train():
    oh_data = pd.get_dummies(train) #对非数值数据进行 ont-hot 编码
    columns_size = oh_data.columns.size
    X_train = oh_data.iloc[:, 1:columns_size-2] #取特征
    X_train = preprocessing.scale(X_train) #归一化
    return X_train

def get_X_test():
    oh_test = pd.get_dummies(test) #对非数值数据进行 ont-hot 编码
    columns_size = oh_test.columns.size
    X_test = oh_test.iloc[:, 1:columns_size-2] #取特征
    X_test= preprocessing.scale(X_test) #归一化
    return X_test
X_test = get_X_test()
X_train = get_X_train()
y_test = get_y_test()
y_train = get_y_train()

print('X_test.shape:'+str(X_test.shape))
print('X_train.shape:'+str(X_train.shape))

from keras.layers import Dense, LSTM, Dropout
from keras.models import Sequential
from keras import optimizers
from sklearn import metrics

model = Sequential()
model.add(Dense(25, input_dim = 25, activation = 'relu'))
model.add(Dropout(0.2))
model.add(Dense(128, activation = 'relu'))

```

```

model.add(Dropout(0.2))
model.add(Dense(64, activation = 'relu'))
model.add(Dense(1, activation = 'sigmoid'))
model.compile(optimizer = optimizers.Adam(lr = 0.001), loss =
'binary_crossentropy', metrics = ['accuracy'])

history = model.fit(X_train, y_train, epochs = 20, batch_size = 2520,
validation_data = (X_test, y_test))

y_pred = model.predict_classes(X_test, batch_size = 10, verbose = 1)
print(y_pred)

target_names = ['1', '0']
print(metrics.classification_report(y_test, y_pred,
target_names = target_names))
df1 = pd.DataFrame(y_pred)
df1
#模型评估
history_dict = history.history
loss, accuracy = model.evaluate(X_test, y_test)

print('test loss: ', loss)
print('test accuracy: ', accuracy)

history_dict = history.history
loss_values = history_dict['loss']
val_loss_values = history_dict['val_loss']
acc = history_dict['accuracy']
epochs = range(1, len(acc) + 1)
print(history_dict)

#画 epoc_acc 图
plt.figure(figsize=(8,6), facecolor='white', edgecolor='white') #设置画布颜色、
画布边线颜色
plt.plot(epochs, acc_values, 'bo', label = 'Training acc')
plt.plot(epochs, val_acc_values, 'b', label = 'Validation acc')
plt.title('Training and validation accuracy')
plt.xlabel('Epochs')
plt.ylabel('Acc')
plt.legend()
plt.show()

#画 epoch_loss 图
plt.figure(figsize=(8,6), facecolor='white', edgecolor='white') #设置画布颜色、

```

画布边线颜色

```
plt.plot(epochs, loss_values, 'bo', label = 'Training loss')
plt.plot(epochs, val_loss_values, 'b', label = 'Validation loss')
plt.title('Training and validation loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.show()
acc_values = history_dict['accuracy']
val_acc_values = history_dict['val_accuracy']
```

#导出预测 result 和实际 result 导出 loss_acc 表导出训练集、测试集

```
Df1= pd.DataFrame(y_pred,y_test)
Df1.to_csv("out_result1.csv")
print(pd.DataFrame(history_dict))
Df2=pd.DataFrame(history_dict)
Df2.to_csv("loss_acc1.csv")
```

```
train1,test1= train_test_split(data,test_size=0.20,random_state=10)
Df3= pd.DataFrame(test1)
Df3.to_csv("out_test.csv")
Df4= pd.DataFrame(train1)
Df4.to_csv("out_train.csv")
```

2、MATLAB 附加动量的 BP 神经网络代码

%% 清空环境变量

```
clc
clear
```

%输入数据归一化

```
input_train=input_train';
output_train=output_train';
input_test=input_test';
[inputn,inputps]=mapminmax(input_train);
```

%% 网络结构初始化

```
innum=25;
midnum=25;
outnum=2;
```

%权值初始化

```

w1=rands(midnum, innum);
b1=rands(midnum, 1);
w2=rands(midnum, outnum);
b2=rands(outnum, 1);

w2_1=w2;w2_2=w2_1;
w1_1=w1;w1_2=w1_1;
b1_1=b1;b1_2=b1_1;
b2_1=b2;b2_2=b2_1;

%学习率
xite=0.1;
alfa=0.01;
loopNumber=10;
I=zeros(1, midnum);
Iout=zeros(1, midnum);
FI=zeros(1, midnum);
dw1=zeros(innum, midnum);
db1=zeros(1, midnum);

%% 网络训练
E=zeros(1, loopNumber);
for ii=1:10
    E(ii)=0;
    for i=1:1:1964
        %% 网络预测输出
        x=inputn(:, i);
        % 隐含层输出
        for j=1:1:midnum
            I(j)=inputn(:, i)'*w1(j, :)' +b1(j);
            Iout(j)=1/(1+exp(-I(j)));
        end
        % 输出层输出
        yn=w2'*Iout'+b2;

        %% 权值阈值修正
        %计算误差
        e=output_train(:, i)-yn;
        E(ii)=E(ii)+sum(abs(e));

        %计算权值变化率
        dw2=e*Iout;
        db2=e';
    end
end

```



```

        for j=1:1:midnum
            S=1/(1+exp(-I(j)));
            FI(j)=S*(1-S);
        end
        for k=1:1:innum
            for j=1:1:midnum
                dw1(k,j)=FI(j)*x(k)*(e(1)*w2(j,1)+e(2)*w2(j,2));
                db1(j)=FI(j)*(e(1)*w2(j,1)+e(2)*w2(j,2));
            end
        end

        w1=w1_1+xite*dw1'+alfa*(w1_1-w1_2);
        b1=b1_1+xite*db1'+alfa*(b1_1-b1_2);
        w2=w2_1+xite*dw2'+alfa*(w2_1-w2_2);
        b2=b2_1+xite*db2'+alfa*(b2_1-b2_2);

        w1_2=w1_1;w1_1=w1;
        w2_2=w2_1;w2_1=w2;
        b1_2=b1_1;b1_1=b1;
        b2_2=b2_1;b2_1=b2;
    end
end

%%分类
inputn_test=mapminmax('apply',input_test,inputps);
fore=zeros(2,150);
for ii=1:1
    for i=1:15%1500
        %隐含层输出
        for j=1:1:midnum
            I(j)=inputn_test(:,i)'*w1(j,:)' +b1(j);
            Iout(j)=1/(1+exp(-I(j)));
        end

        fore(:,i)=w2'*Iout'+b2;
    end
end

%% 结果分析
%根据网络输出找出数据属于哪类
output_fore=zeros(1,15);
for i=1:15

```

```
    output_fore(i)=find( fore(:,i)==max( fore(:,i)) );  
end
```