

# 融合历史多维数据的油品销售模型构建及应用实践

赵 亮（中国石油化工股份有限公司云南石油分公司，云南 昆明 650011）

**摘 要：**制定营销策略是加油站业务的一个重要组成部分，而数据驱动的营销策略制定已经变成了加油站实现精准营销的迫切需要。因此，我们提出了一种以加油站历史数据、营销策略和关键特征为基础的油品销量预测的深度学习模型，以及以销量预测模型为基础的营销策略制定方法。以加油站历史数据特点为基础，设计了一个多层次的网络结构，对不同类型特征的数据进行处理，并与营销策略信息相结合，来进行油品的销量预测。在此基础上，对加油站的营销策略进行修改，使其能够自动地进行营销策略的选择。通过对实际加油站数据建立的数据集进行试验，结果表明：与其他主流方法相比，本项目提出的销售预测模型的预测误差更小。

**关键词：**历史数据；销售模型；深度学习

## 0 引言

针对当前加油站不同区域、不同特点的情况下，如何制定并执行不同的营销策略（即精准营销）的问题，本项目拟采用机器学习的方法，对其在不同区域、不同特点的加油站的历史数据进行训练，从而使其能够针对不同区域、不同特点的加油站，制定并执行个性化的营销策略（简称“精准营销”）。基于数据驱动的不同市场策略下的销售预测能够为企业的销售决策提供参考。因此，本项目提出了一种基于数据的加油站销售预测方法。

## 1 相关工作

### 1.1 销量预测问题

企业为了获得更大的利润，在销售预测、重要信息的维护等方面，一直都在不断地进行改进。由于销售预测工作所需的资料多为时序资料，因此可用时序模式如 ARIMA 来进行销售预测。此外，广义线性模型，决策树，梯度提升树等机器学习方法也被广泛用于销售预测。在对历史数据进行预测时，研究人员可以根据时间窗的形式，抽取其中的特征，然后利用机器学习的方法对其进行预测。通过对销售量的预测，可以确定销售量的影响因素，从而为制订营销战略提供依据。同时，深度学习技术也被用于销售预测。一些学者利用循环神经网络和注意机制对关联的时间序列进行了分析，并将其应用于汽车销售量的预测。

### 1.2 时间序列模型

基于时序特征的高维特征，如 SVM，LR，决策树，Adaboost 等，造成了资源消耗大，效率低下等问题。为了解决这一问题，我们提出了循环神经网络及其两个著名的变种：长度记忆网络（LSTM）、门控循环单元（GRU）。对于时序数据，RNN 和它的变种都表现

出了良好的性能；另外，基于注意的思想，我们提出了一种新的方法来给时间序列数据赋权，从而使深度学习能够更好地将重点放在时间序列数据上。之后，我们又建立了一个独立于循环神经网络的自注意机制，并将其应用于时间序列数据，得到了良好的结果。

## 2 加油站销量预测模型与方法

在这一部分，首先介绍了要解决的问题，包括与此问题有关的与数据预处理有关的工作。进而，本项目将构建一套面向加油站销售的深度学习模型，并将其应用于加油站销售预测。

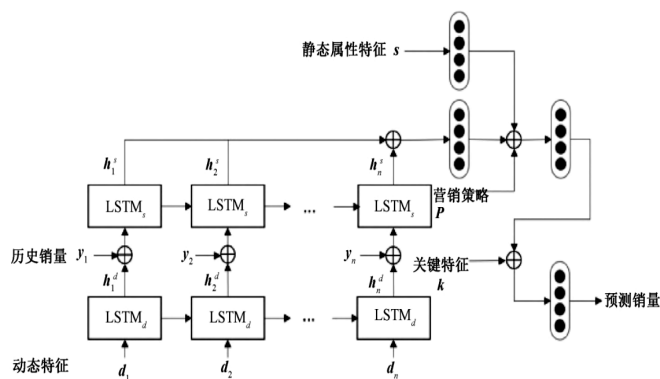


图1 基于深度学习的油品销量预测模型框架

所输入的数据主要有：加油站的静态特性，如占地面积，油枪数量等；加油站在过去  $n$  d 动态特性  $D=\{d_1, d_2, \dots, d_n\}$ ， $d_i$  包含历史行销策略，节日，天气等；一个加油站在过去  $n$  d 的销售数据  $Y=\{y_1, y_2, \dots, y_n\}$ ；在第  $j$  日，加油站打算采用的行销策略  $p_j$ ；第  $j$  日的主要特点资料  $k_j$ 。结果为预测的油站原油销售量  $\hat{y}_i$ ，其计算公式为：

$$\hat{y}_i = F(S, D, Y, p_j, k_j), j \geq n+1\#$$

(1)

其中： $F(\cdot)$ 是一个非线性的预报函数，它是预报模型的一个数学表达。

2.1 数据预处理

在实际应用中，与加油站营销有关的原始数据（如文字数据）形式多样，而在模型的训练和测试中，对数值型数据的需求较大。针对燃油服务中文字型数据大多属于枚举型，在数据预处理时，对每一种枚举型的特性建立一个数值型的编码列表，并将这些特性转换成数字型。当各参数之间没有尺寸相关时，则采用独立热码对原参数进行编码。但是，对于取值类型为连续型数值变量的原始特征，为了排除不同特征存在的不同尺度取值范围带来的影响，使用了归一化的方法，将原始数值转化为  $[0, 1]$  中的数值。

2.2 数据特征表示

为了解决这一时间特性，我们采用了一种基于双 LSTM 的网络结构。设  $LSTM_{dyn}$  为一网络，其对一历史动态数据进行处理，给定输入  $d_i$ ，每个周期单位将生成一隐含的输出  $h_i^d$ ；令  $LSTM_{sal}$  为处理历史销售量数据的网络，以接收拼接后的  $h_i^d$  和  $y_i$  为输入的每个循环单元，输出综合了历史动态信息和历史销售量信息的隐藏层表示  $h_i^s$ 。

$$h_i^d = LSTM_{dyn}(h_{i-1}^d, d_i) \quad (2)$$

$$h_i^s = LSTM_{sal}(h_{i-1}^s, h_i^d, y_i) \quad (3)$$

最后，利用特征抽取模块，对加油站静态数据、历史动态数据和历史销售量数据进行了隐含层表式  $s_i$ ， $\{h_i^s\} 1 \leq i \leq n$ ，并将其输入到后续的销售量预测网络中。

3 实验与结果

3.1 数据集

通过对 201 家加油站在 2020 年 1 月 1 日至 12 月 31 日期间的销售情况进行分析，建立了销售预测模型。数据集集中的每个数据包括静态特性，动态特性，市场策略和销售数据。在试验中，由于实际情况下的销售量是根据各个站点的历史数据来对将来的销售量进行预测，因此，我们采用从 3 月 1 日到 10 月 31 日的时间间隔作为训练数据集（由于新冠疫情的影响，我们去掉了 2020 年 1-2 个月的时间间隔），从 11 月 1 日到 12 月 31 日的时间间隔作为测试数据集。在试验中，设定了 30 天的加油站历史数据，并建立了相应的模型，

对未来一到七天的销售量进行了预测。

3.2 评价指标

结合工程实例，采用了两个评估指标对该模型进行了评估。一种直觉的评估指数为平均绝对误差（MAE），其计算公式由（4）表示。另外，在考虑到不同地点的销售规模的情况下，一个替代性的评估指标是平均相对误差（MRE），其计算公式如下：

$$MER = \frac{1}{m} \sum_{k=1}^m \frac{|y_j^k - \hat{y}_j^k|}{y_j^k} \quad (4)$$

在实际运用中，与均方误差相比，平均绝对误差、平均相对误差能更好地反映出模型的预报效果。

3.3 单日销量预测实验

在建立的油品销售预测数据集上，采用多种方法对油品销售数据进行了分析。

表 1 预测未来第 1 天销量的实验结果

| 方法               | MAE    | MRE    |
|------------------|--------|--------|
| K 近邻             | 0.7824 | 0.1248 |
| 极端随机树            | 0.8165 | 0.1384 |
| Adaboost         | 0.8518 | 0.1351 |
| GradientBoosting | 0.8146 | 0.1324 |
| 基于关键特征的模型        | 0.6484 | 0.1061 |

试验表明，与其他常见的预测方法相比，该方法具有最小的预测误差。本项目拟采用“端到端”的方法，对各类原始特征进行分类，对网络进行训练，调整各特征的权重，并在网络中引入关键特征，使其具有较高的预测精度。

3.4 模型结构对销量影响实验

通过对三种模型的分析，分别对三种模型进行了比较，第一种模型去掉了主要的特征项；第 2 版基于第 1 版，对拼接后的销售数据及动力学特性数据只采

用了一个 LSTM 网络进行处理;第 3 版基于前一版,只将 LSTM 中最后一次周期单元的输出值,而没有将全部周期单元的输出值用于预测网络。表 2 列出了 3 种不同的模型及原模型(没有重要特性)的试验结果。

表 2 不同版本模型销量预测的实验结果

| 模型版本         | MRE    |
|--------------|--------|
| 无关键特征        | 0.1853 |
| 无关键特征和单 LSTM | 0.2283 |
| 无关键特征和单个单元   | 0.3517 |
| 基于关键特征的模型    | 0.1068 |

通过对模型的分析,验证了该方法的有效性。

### 3.5 销量预测可视化实验

我们选取了 3 家加油站作为研究对象,利用它们的预测值与实测值,对它们进行了比较,得到了它们之间的关系曲线。图 2 显示了销售预测。试验表明,该模型的销售预测值与实际销售值之间的差异很小,并且预测值与实际值的趋势是一致的。因而,本研究之预测模式,既有良好之预测精度,又能捕捉到销售变动之趋势。

## 4 结论

本项目面向油品销售预测导向下的精准营销需求,研究面向加油站销售终端的油品销售预测与营销策略优选方法。本项目利用有目标的特征抽取网络架构,从不同类型加油站的历史数据中提取出影响销售量的因素,并将其与市场营销决策相结合,实现销售量的预测。为了满足市场的迅速变化以及营销战略的敏感度等需要,本文使用了四种相关性分析方法,将与销售量有较高关联度的关键特征提取出来,并将它们加入到预测网络中,从而有效地减少了模型的预测误差。实践者藉由观察不同行销策略之销售产出,来决定行销策略之成效。通过实例分析,验证了该方法的正确性。在此基础上,我们还将结合现实中可能存

在的一些问题(如节假日、突发事件等),以及可能对销售预测产生影响的因素,在此基础上,根据模型应用的实际情况,对油站销售预测模型进行持续的优化。

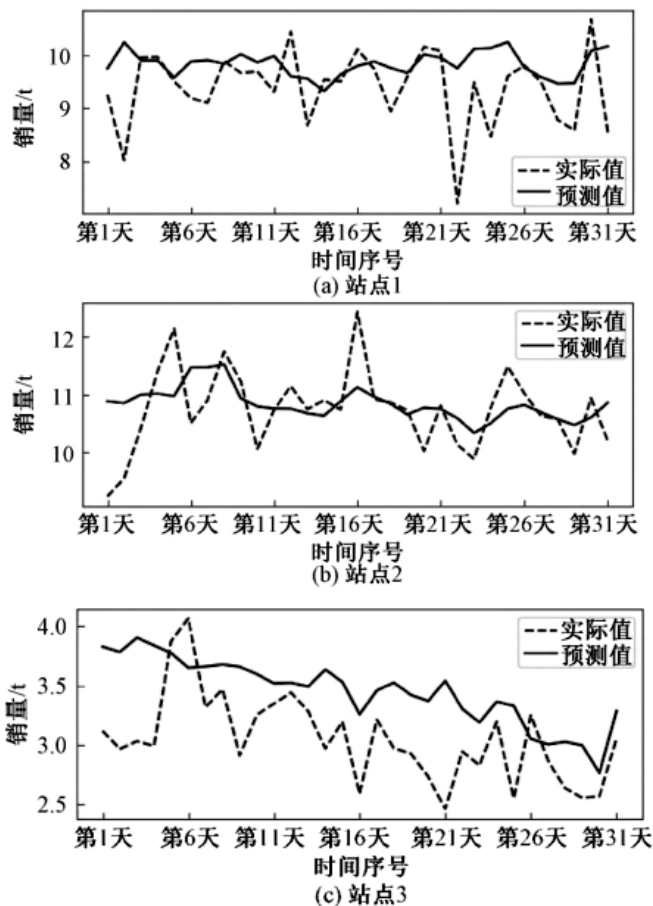


图 2 基于关键特征的销量预测模型结构图

### 参考文献:

- [1] 孙岩,雷超,董恒,等.历史数据下优化 GRU 模型的测控设备相位预测[J].测控技术,2023,42(1):16-20.
- [2] 沈永发,郑煜,刘辉冉,等.基于历史数据的轨道交通站内客流预测模型[J].武汉理工大学学报:信息与管理工程版,2022,44(6):887-893.
- [3] 何源,何亮.白酒企业新产品销售预测模型构建研究[J].酿酒,2023,50(2):29-31.
- [4] 陈世楼,洪松琛,孙鹏.基于历史数据分析的采购策略智能推荐模型研究[J].企业改革与管理,2021(9):2.

### 作者简介:

赵亮(1973-),男,汉族,云南昆明人,本科,高级工程师,研究方向:信息和数字化转型有关项目。