

Precise Demand Forecast for New Retail Target Products

Yanqin Zhang, Xiangfang Yan, Zhifei Liu, Hong Lan*

Jiangxi University of Science and Technology, Ganzhou Jiangxi
Email: *40376334@qq.com

Received: Jul. 30th, 2020; accepted: Aug. 14th, 2020; published: Aug. 21st, 2020

Abstract

This paper clarifies the three important factors that affect the sales volume of commodities, constructs the gray correlation model and the gray forecast GM(1,1) model, and uses big data analysis technology according to the historical sales data of a certain product to accurately predict the sales volume of the product. It can effectively solve the current problems that most new retail companies focus on.

Keywords

Sales Forecast, Big Data Analysis, Grey Forecast GM(1,1) Model

新零售目标产品的精准需求预测

张燕琴, 严祥方, 刘志飞, 兰红*

江西理工大学, 江西 赣州
Email: *40376334@qq.com

收稿日期: 2020年7月30日; 录用日期: 2020年8月14日; 发布日期: 2020年8月21日

摘要

本文通过明确影响商品销售量三个重要因素、构建灰色关联模型和灰色预测GM(1,1)模型, 同时根据某商品历史销售数据采用大数据分析技术, 对商品的销售量进行了精准预测, 可有效解决当前大多数新零售企业重点关注的问题。

关键词

销售预测, 大数据分析, 灰色预测GM(1,1)模型

*通讯作者。

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着我国消费市场的不断发展, 市场上的消费模式已经逐步由“以物为主”转变为“以客为主”。在新零售行业, 性价比不再是顾客衡量是否购买物品的唯一标准, 人们的需求也不仅仅是单一的追求实用性, 而是更多的考虑时尚性, 把注意力放在“个性化、时尚、美观”等方面。在这类特殊需求的推动下, 新零售企业的生产模式逐步向多品种、小批量迈进, 这让商场内零售店铺里的饰品和玩具等种类变得更加琳琅满目, 同时也给零售行业的库存管理增加了很大的难度。如何根据层级复杂, 品类繁多的历史销售数据, 以区域层级, 小类层级乃至门店 skc (单款单色) 层级给出精准的需求预测, 是当前大多数新零售企业需要重点关注并思考的问题。

2. 基于灰色关联分析的模型

考虑到各种因素对目标商品销售量的影响, 结合实际生活和附件信息, 选取了三个因素进行研究, 分别是库存信息、节假日折扣、预计加权销售额。需要确定累计销售额排名前 50 的商品, 然后可根据近大远小的加权原则[1]分别预测得四个节日的销售额, 将此三个作为影响销售量的因子, 采用灰色关联度分析得到各因素对 skc 销售量的影响情况。

2.1. 计算流程

本问首先需要确定从 2018 年 7 月 1 日至 2018 年 10 月 1 日累计销售额最多的 skc, 由于附件数据非常庞大, 使用大数据的方法进行求解, 求解思路如图 1 所示。

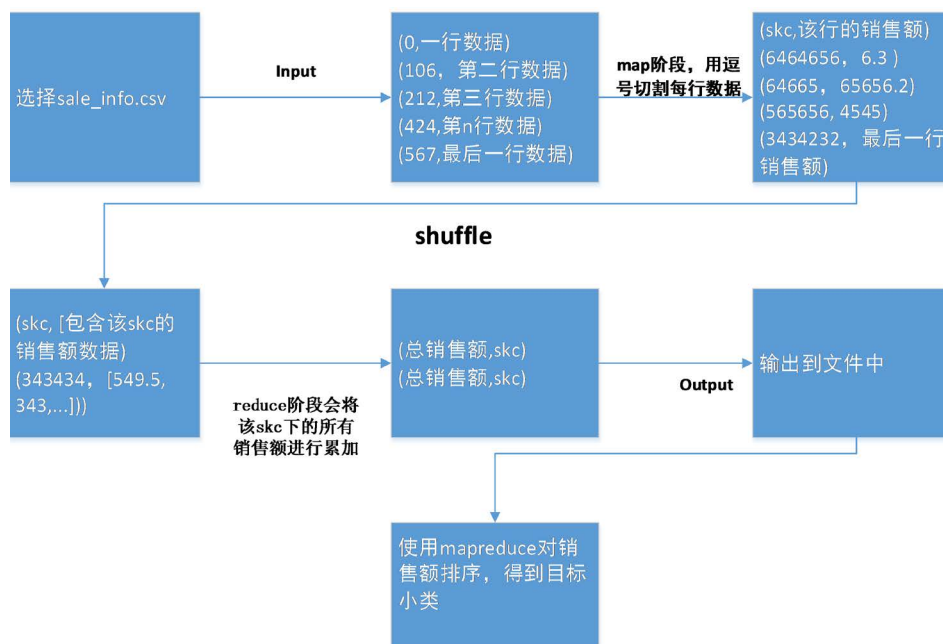


Figure 1. Flow chart of big data solving target skc

图 1. 大数据求解目标 skc 流程图

利用大数据框架 Hadoop 的 mapreduce 进行数据处理, 对 sale_info.csv 的数据, 在 map 阶段对数据切割, 因为 csv 格式是按逗号分隔的, 用 Java 的 split 函数以逗号分隔。其次把 skc 编号作为 key, 把销售额当成 value, 同一个 key 会对应到同一个 reduce 中, 对同一个 key 的 value 即销售额进行累加得到一个 sum 值, 再将 sum 作为 key, skc 编号作为 value, 即可算出该 skc 的销售额。最后将算出的数据再次进行 mapreduce, 自定义排序方法为降序排序, 把 sum 值当成 key。取销售额最多的 50 个, 即为目标 skc。

2.2. 灰色关联分析模型的建立

在灰色系统理论[2]中, 提出了一种新的关联度分析方法, 即根据因素之间发展态势的相似或相异程度来衡量因素间关联的程度, 它揭示了事物动态关联的特征与程度。

1) 参考数列的选取

选择某节日的目标商品你的销售量 x_0^i 为参考数列, 则有如下表达式。

$$x_0^i = \{x_0^i(k) | k = 1, 2, 3, \dots, n\} = \{x_0^i(1), x_0^i(2), \dots, x_0^i(n)\} \quad (1)$$

其中, i 表示节日, i 为 0 时为国庆节, i 为 1 时表示双十一, i 为 2 时表示双十二, k 表示时刻。

2) 比较数列的选取

选择影响目标 skc 的变量 x_j^i 为比较数列, 则有如下式子。

$$x_j^i = \{x_j^i(k) | k = 1, 2, \dots, n\} = \{x_j^i(1), x_j^i(2), \dots, x_j^i(n)\} \quad (2)$$

其中, $j = 1$ 时表示为第 i 个节日的库存信息, $j = 2$ 时表示为第 i 个节日的折扣, $j = 3$ 时表示为第 i 个节日的加权预计销售额。

3) 关联度的建立

根据灰色关联分析, 得到比较数列 x_j^i 对参考数列 x_0^i 在 k 时刻的关联系数 λ_j^i , 其表达式如下。

$$\lambda_j^i = \frac{\min(s) \min(t) |x_0^i(t) - x_j^i(t)| + \rho \max(j) \max |x_0^i(t) - x_j^i(t)|}{|x_0^i(k) - x_j^i(k)| + \rho \max(j) \max |x_0^i(k) - x_j^i(k)|} \quad (3)$$

其中, ρ 为分辨率系数, ρ 越大, 分辨率越高; ρ 越小, 分辨率越高; 为了将(3)式中的各个时刻的分散的关联数集中分析, 因此在它的基础上, 定义一个新的式子表示 x_j^i 对参考数列 x_0^i 的关联度为 r_j^i

$$r_j^i = \frac{1}{n} \sum_{k=1}^n \lambda_j^i(k) \quad (4)$$

2.3. 模型的求解

代入已计算得出的结果, 结果如表 1 所示。

Table 1. Correlation value

表 1. 关联度值

| 节日 | 库存关联度 | 折扣关联度 | 加权预计销售额关联度 |
|-----|-------|-------|------------|
| 国庆 | 0.666 | 0.590 | 0.730 |
| 双十一 | 0.694 | 0.712 | 0.751 |
| 双十二 | 0.644 | 0.701 | 0.774 |

总体来看, 在这三个因素中, 加权预计销售额的相关程度都是最高的, 可见它对目标 skc 的影响颇大, 而在双十一、双十二时期, 库存相关程度都是最低的, 这与实际情况相符合, 现实生活中, 双十一、

双十二只有一天，因此商家都会预先准备充分的产品供用户购买，因此相对来说，它库存的相关程度是最低的，而在国庆时期，由于是全国统一的节假日，而且有七天的时间，用户购买欲较平常而言相对较高，而商家没有办法一次性准备充分的产品，因此货存相对而言它的相关程度就会比折扣更高。

3. 灰色预测 GM(1,1)模型的建立与求解

灰色预测是指利用 GM 模型对系统行为特征的发展变化规律进行估计预测，同时也可以对行为特征的异常情况发生的时刻进行估计计算，以及对在特定时区内发生事件的未来时间分布情况做出研究等等。这些工作实质上是将“随机过程”当作“灰色过程”，“随机变量”当作“灰变量”，并主要以灰色系统理论中的 GM(1,1)模型[3]来进行处理。

3.1. 数据准备

本问首先需要确定从 2019 年 6 月 1 日至 2019 年 10 月 1 日累计销售额最多的小类，由于附件数据非常庞大，使用大数据的方法进行求解，求解思路如图 2 所示。

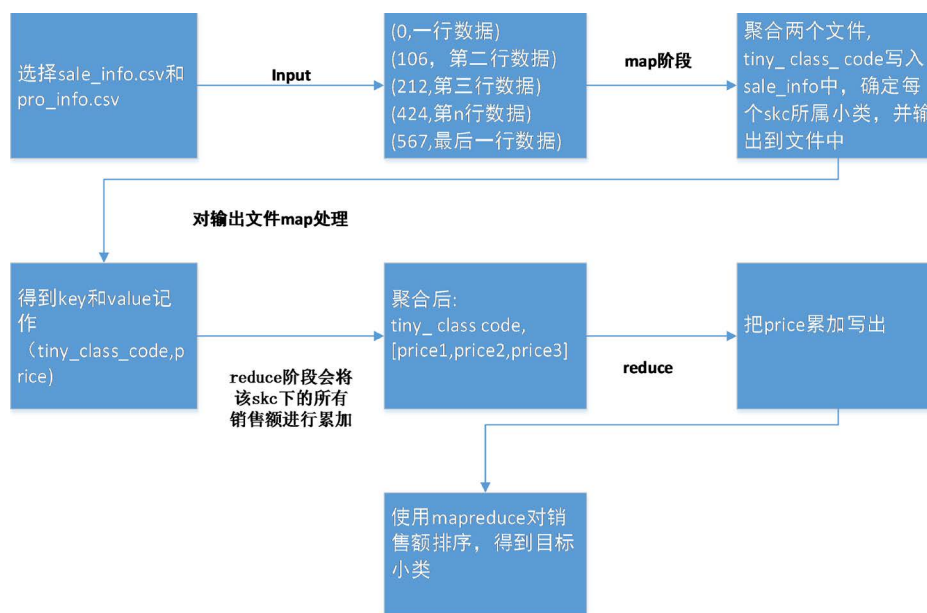


Figure 2. Flow chart of solving target subcategories of big data
图 2. 大数据求解目标小类流程图

由图可知，确定目标小类时需要使用附件 sale_info.csv 和 prod_info.csv 的数据，首先对其进行一次 map 聚合，把 prod_info 附件中的 tiny_class_code 属性写入到 sale_info.csv 的与之相对应的 skc，以此确定每个 skc 所属小类。其次对 sale_info 的数据，使用大数据技术，在 mapreduce 和 map 阶段，对一行数据切割，取 tiny_class_code 作 key，取销售额为 value，同样写入 reduce，同一个 key 写入同一个 reduce，将销售额进行累加，得到每个小类的销售额。最后把 tiny_class_code 为 key，总销售额为 value，即可得出 21 个小类在 2019-06-01 到 2019-10-01 中的销售额，选出前十的小类即可。

3.2. 灰色预测 GM(1,1)模型

1) 参考数列的选择

本问主要研究的是预测问题，即预测目标小类在 2019 年 10 月 1 日后 3 个月中每个月的销售量，因

此选择各小类的 1 到 9 月的销售量作为参考数列 $y_m(0)$ ，其中

$$y_m(0) = (y_m^1(0), y_m^2(0), \dots, y_m^n(0)) \quad (5)$$

m 表示第几个小类， n 表示第几个月。

2) 累加数列与均值数列的选择

在式(5)的基础上，做一次累加(AGO)得到累加数列 $y_m(1)$ ，则有

$$y_m(1) = (y_m^1(0), y_m^1(0) + y_m^2(0), \dots, y_m^{n-1}(0) + y_m^n(0)) \quad (6)$$

在此基础上，得到均值数列

$$z_m^n(1) = 0.5y_m^n(1) + 0.5y_m^{n-1}(1), n = 2, \dots, 9 \quad (7)$$

3) GM(1,1)模型的建立

基于参考数列、累加数列和均值数列建立得到灰微分方程为

$$y_m^n(0) + az_m^n(1) = b, n = 2, 3, \dots, 9 \quad (8)$$

由最小二乘法求得达到最小值的 $\hat{u} = (a, b)^T = (B^T B)^{-1} B^T Y$ ，因此可以求得(1,1)模型为

$$y_m^{n+1}(1) = \left(y_m^1(0) - \frac{b}{a} \right) e^{-an} + \frac{b}{a}, k = 1, 2, \dots, 8 \quad (9)$$

3.3. GM(1,1)模型的求解

为了保证模型的可靠性，需要对数据进行预处理和检验。

1) 求解步骤

首先计算参考数列 $y_m(0) = (y_m^1(0), y_m^2(0), \dots, y_m^n(0))$ 的级比，级比数列定义为

$$J_m^n = \frac{J_m^{n-1}(0)}{J_m^n(0)}, n = 2, 3, \dots, 9 \quad (10)$$

在计算时需要所有的 J_m^n 的范围都处于 $\left(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}} \right)$ 中，以保证可以使用参考数列 $y_m(0)$ 作为处理灰色预测模型 GM(1,1) 的数据。当 J_m^n 的范围不处于 $\left(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}} \right)$ 之中时，取适当的常数 c ，作平移得到新的参考数列

$$s_m(0) = y_m(0) + c \quad (11)$$

$s_m(0) = (s_m^1(0), s_m^2(0), \dots, s_m^n(0))$ ，与原参考数列的关系为从而也得到一个新的级比 Q_m^n 其定义为

$$Q_m^n = \frac{Q_m^{n-1}(0)}{Q_m^n(0)}, n = 2, 3, \dots, 9 \quad (12)$$

根据式(10)即可求得在 10 月、11 月和 12 月的预测值。

2) 数据的检验

通过上述方式预测出的销售量可能存在较大误差，因此需要做检验，从两个角度进行检验，第一个是残差检验，第二是级比偏差值检验。

令残差为 $\alpha_m^n(0) = \frac{y_m^n(0) - \hat{y}_m^n(0)}{y_m^n(0)}, n = 1, 2, \dots, 9$ ，若残差 $\alpha_m^n(0) < 0.2$ ，则可认为达到一般要求，若残差

$\alpha_m^n(0) < 0.1$ ，则可认为达到较高要求。

根据级比数列 $J_m^n = \frac{J_m^{n-1}(0)}{J_m^n(0)}$, $n = 2, 3, \dots, 9$ ，结合发展系数 a 求出相应的级比偏差 $\beta_m^n = 1 - \left(\frac{1-0.5a}{1+0.5a} \right) J_m^n$ ，

若残差 $\beta_m^n < 0.2$ ，则可认为达到一般要求，若残差 $\beta_m^n < 0.1$ ，则可认为达到较高要求。

3.4. 模型结果

通过编写 c 语言程序进行求解，得到目标小类在 2019 年 10 月、11 月和 12 月的销售量的预测值，部分计算结果如表 2 所示。根据所求数据使用计算得出各个月份的 MAPE 值如表 2 所示。

Table 2. Monthly MAPE

表 2. 每月 MAPE

| 月份 | MAPE |
|------|-------|
| 10 月 | 0.089 |
| 11 月 | 0.154 |
| 12 月 | 0.160 |

每个月的 MAPE 值均小于 0.17，因此可以认为使用此模型进行预测，其准确度较高，可靠性较强。

4. 引入平衡因子的优化 GM(1,1)模型

4.1. 数据准备

在问题二的基础上，继续找到每个类的所有 skc，为了更精准的预测 skc 每月的周销量，需要计算得到每个 skc 在 2019 年每个月内每周的周销量，同样使用大数据分析，求解思路如图 3 所示。

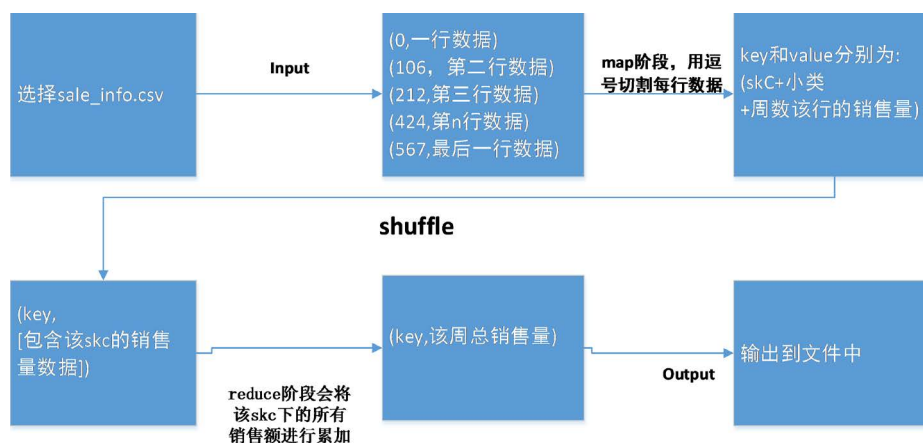


Figure 3. Weekly sales flow chart for solving skc with big data

图 3. 大数据求解 skc 的周销量流程图

根据 tiny_class_code 的 sale_info，用 Java 的 Date 类进行 Date 化，接着对 Date 对象 Calendar 化，取出该周是 2019 年的第几周，先把 2019-09-30 的数据对应第 40 周，因此取第四十周为第一周，取 30~51 周的数据，即前十周和后 12 周数据。取出该时间段数据后，把 key 写成 skc,tiny_class_code,week，把 value 写成销售量，写出到 reduce，同样对 value 相加，sum 为 value，得到相应的 value 值，即可得出前十周和后十二周的数据。

4.2. 优化后的 GM(1,1)模型

根据本问的要求以及计算结果,依次得到累计销售额排名前十的小类对应的 skc 分别有 11417、1676、3777、1475、6125、5906、2362、634、1257、721 个,数据量非常大,并且需要预测每周的销售量。为了提高数据的精准性和可靠性,对各类商品的需求进行精准预测,因此可在问题二所建立的模型基础上,引入平滑因子 c [4],得到一个新的模型,称之为 BGM(1,1)模型。

在问题二的模型中,有参考数列 $y_m(0) = (y_m^1(0), y_m^2(0), \dots, y_m^n(0))$,对它做 1-AGO 操作,得到 $y_m(1) = (y_m^1(0), y_m^1(0) + y_m^2(0), \dots, y_m^{n-1}(0) + y_m^n(0))$,为了满足一定的平滑条件,则需要满足以下表达式

$$\begin{cases} 0 \leq \frac{y_m^n(0)}{y_m^n(1)} \leq 0.5 \\ 0 \leq \frac{y_m^n(0)}{y_m^{n-1}(1)} \leq 1.5 \end{cases} \quad (13)$$

在引入平滑因子 c 之后,可得

$$\begin{cases} y_m^n(1) = y_m^n(0) + c \\ y_m^n(1) = \sum_{i=1}^n y_m^i(0) \end{cases} \quad (14)$$

结合式(15)和式(16),可得

$$\begin{cases} c \geq 0 \\ c \geq \max \left\{ \frac{2y_m^n(0) - y_m^{n-1}(1)}{n-3} \right\} \end{cases} \quad (15)$$

4.3. 模型的结果

使用问题二的求解步骤,并增加约束条件式(17),编写 C 语言程序,求解得到以下结果,通过编写 c 语言程序,求得 2019 年 10 月 1 日后 12 周的预测 MAPE。如表 3 所示。

Table 3. Weekly MAPE
表 3. 每周 MAPE

| 周数 | MAPE | 周数 | MAPE |
|----|-------|----|-------|
| 1 | 0.172 | 7 | 0.221 |
| 2 | 0.162 | 8 | 0.225 |
| 3 | 0.140 | 9 | 0.243 |
| 4 | 0.115 | 10 | 0.227 |
| 5 | 0.215 | 11 | 0.145 |
| 6 | 0.177 | 12 | 0.184 |

总体来看,每周的 MAPE 维持在一定范围上,但均小于 0.25,在周销量信息如此不足的情况下,预测效果总体来看可以认为是可靠的。

5. 结束语

通过 Excel、Java 语言等工具完成数据的筛选及分析,建立了影响商品销售量及销售额的灰色关联模型,其中最相关因素分别为库存信息、节假日折扣、预计加权销售额,并在此基础上建立了灰色预测

GM(1,1)模型以及在引入平滑因子后的优化模型。本模型综合考虑了商品库存、节假日等实际因素，在一定程度上能精准预测商品的销售量及销售额。可作为零售企业对各类商品的需求量的参考。

参考文献

- [1] 谢东闻. “Excel 表”在销售预测中的应用[J]. 会计研究, 2017(13): 91-92.
- [2] 章绍辉. 数学建模[M]. 北京: 科学出版社, 2010: 416-419.
- [3] 王少然, 刘文慧. 基于 GM(1,1)回归的需求预测优化模型研究[J]. 慧眼聚焦, 2016(11): 1-4.
- [4] 黄鸿云, 刘卫校, 丁佐华. 基于多维灰色模型及神经网络的销售预测[J]. 软件学报, 2019, 30(4): 10.